AUTHENTICATED
U.S. GOVERNMENT
INFORMATION
GPO

BEHAVIORAL SCIENCE AND SECURITY: EVALUATING TSA'S SPOT PROGRAM

HEARING

BEFORE THE

SUBCOMMITTEE ON INVESTIGATIONS AND OVERSIGHT

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

HOUSE OF REPRESENTATIVES

ONE HUNDRED TWELFTH CONGRESS

FIRST SESSION

APRIL 6, 2011

Serial No. 112–11

Printed for the use of the Committee on Science, Space, and Technology

Available via the World Wide Web: http://science.house.gov

U.S. GOVERNMENT PRINTING OFFICE

65–053PDF WASHINGTON : 2011

For sale by the Superintendent of Documents, U.S. Government Printing Office
Internet: bookstore.gpo.gov Phone: toll free (866) 512–1800; DC area (202) 512–1800
Fax: (202) 512–2104 Mail: Stop IDCC, Washington, DC 20402–0001

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

HON. RALPH M. HALL, Texas, *Chair*

F. JAMES SENSENBRENNER, JR., Wisconsin
LAMAR S. SMITH, Texas
DANA ROHRABACHER, California
ROSCOE G. BARTLETT, Maryland
FRANK D. LUCAS, Oklahoma
JUDY BIGGERT, Illinois
W. TODD AKIN, Missouri
RANDY NEUGEBAUER, Texas
MICHAEL T. McCAUL, Texas
PAUL C. BROUN, Georgia
SANDY ADAMS, Florida
BENJAMIN QUAYLE, Arizona
CHARLES J. "CHUCK" FLEISCHMANN, Tennessee
E. SCOTT RIGELL, Virginia
STEVEN M. PALAZZO, Mississippi
MO BROOKS, Alabama
ANDY HARRIS, Maryland
RANDY HULTGREN, Illinois
CHIP CRAVAACK, Minnesota
LARRY BUCSHON, Indiana
DAN BENISHEK, Michigan
VACANCY

EDDIE BERNICE JOHNSON, Texas
JERRY F. COSTELLO, Illinois
LYNN C. WOOLSEY, California
ZOE LOFGREN, California
DAVID WU, Oregon
BRAD MILLER, North Carolina
DANIEL LIPINSKI, Illinois
GABRIELLE GIFFORDS, Arizona
DONNA F. EDWARDS, Maryland
MARCIA L. FUDGE, Ohio
BEN R. LUJÁN, New Mexico
PAUL D. TONKO, New York
JERRY McNERNEY, California
JOHN P. SARBANES, Maryland
TERRI A. SEWELL, Alabama
FREDERICA S. WILSON, Florida
HANSEN CLARKE, Michigan

SUBCOMMITTEE ON INVESTIGATIONS AND OVERSIGHT

HON. PAUL C. BROUN, Georgia, *Chair*

F. JAMES SENSENBRENNER, JR., Wisconsin
SANDY ADAMS, Florida
RANDY HULTGREN, Illinois
LARRY BUCSHON, Indiana
DAN BENISHEK, Michigan
VACANCY
RALPH M. HALL, Texas

DONNA F. EDWARDS, Maryland
ZOE LOFGREN, California
BRAD MILLER, North Carolina
JERRY McNERNEY, California

EDDIE BERNICE JOHNSON, Texas

CONTENTS

Date of Hearing

	Page
Witness List	2
Hearing Charter	3

Opening Statements

Statement by Representative Paul C. Broun, Chairman, Subcommittee on Investigations and Oversight, Committee on Science, Space, and Technology, U.S. House of Representatives	16
Written Statement	17
Statement by Representative Donna F. Edwards, Ranking Minority Member, Subcommittee on Investigations and Oversight, Committee on Science, Space, and Technology, U.S. House of Representatives	18
Written Statement	20

Witnesses:

Mr. Stephen Lord, Director, Homeland Security and Justice Issues, Government Accountability Office	
Oral Statement	24
Written Statement	26
Mr. Larry Willis, Program Manager, Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security	
Oral Statement	39
Written Statement	40
Peter J. DiDomenica, Lieutenant Detective, Boston University Police	
Oral Statement	42
Written Statement	44
Dr. Paul Ekman, Professor Emeritus of Psychology, University of California, San Francisco, and President and Founder, Paul Ekman Group, LLC	
Oral Statement	48
Written Statement	50
Dr. Maria Hartwig, Associate Professor, Department of Psychology, John Jay College of Criminal Justice	
Oral Statement	70
Written Statement	71
Dr. Philip Rubin, Chief Executive Officer, Haskins Laboratories	
Oral Statement	79
Written Statement	80

Appendix I: Answers to Post-Hearing Questions

Mr. Stephen Lord, Director, Homeland Security and Justice Issues, Government Accountability Office	114
Mr. Larry Willis, Program Manager, Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security	118
Dr. Paul Ekman, Professor Emeritus of Psychology, University of California, San Francisco, and President and Founder, Paul Ekman Group, LLC	127

Page

Dr. Maria Hartwig, Associate Professor, Department of Psychology, John Jay College of Criminal Justice 130
Dr. Philip Rubin, Chief Executive Officer, Haskins Laboratories 131
Peter J. DiDomenica, Lieutenant Detective, Boston University Police 134

Appendix II: Additional Materials Submitted for the Record

Mr. Stephen Lord, Director, Homeland Security and Justice Issues, Government Accountability Office 140

BEHAVIORAL SCIENCE AND SECURITY: EVALUATING TSA'S SPOT PROGRAM

WEDNESDAY, APRIL 6, 2011

HOUSE OF REPRESENTATIVES,
SUBCOMMITTEE ON INVESTIGATIONS AND OVERSIGHT,
COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY,
Washington, DC.

The Subcommittee met, pursuant to call, at 10:03 a.m., in Room 2318 of the Rayburn House Office Building, Hon. Paul C. Broun [Chairman of the Subcommittee] presiding.

RALPH M. HALL, TEXAS
CHAIRMAN

EDDIE BERNICE JOHNSON, TEXAS
RANKING MEMBER

U.S. HOUSE OF REPRESENTATIVES

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

2321 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6301
(202) 225-6371
www.science.house.gov

Subcommittee on Investigations and Oversight
Behavioral Science and Security: Evaluating TSA's SPOT Program
Wednesday, April 6, 2011
10:00 a.m.-12:00 p.m.
2318 Rayburn House Office Building

Witnesses

Mr. Stephen Lord
Director, Homeland Security and Justice Issues, Government Accountability Office

Transportation Security Administration (Invited)

Mr. Larry Willis
Program Manager, Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security

Dr. Paul Ekman
Professor Emeritus of Psychology, University of California, San Francisco, and President and Founder, Paul Ekman Group, LLC

Dr. Maria Hartwig
Associate Professor, Department of Psychology, John Jay College of Criminal Justice

Dr. Philip Rubin
Chief Executive Officer, Haskins Laboratories

Peter J. DiDomenica
Lieutenant Detective, Boston University Police

HEARING CHARTER

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY
SUBCOMMITTEE ON INVESTIGATIONS & OVERSIGHT
U.S. HOUSE OF REPRESENTATIVES

Behavioral Science and Security: Evaluating TSA's SPOT Program

WEDNESDAY, APRIL 6, 2011
10:00 A.M.—12:00 P.M.
2318 RAYBURN HOUSE OFFICE BUILDING

Purpose

The Subcommittee on Investigations and Oversight meets on April 6, 2011 to examine the Transportation Security Administration's (TSA) efforts to incorporate behavioral science into its transportation security architecture. The Department of Homeland Security (DHS) has been criticized for failing to scientifically validate the Screening of Passengers by Observational Techniques (SPOT) program before operationally deploying it. SPOT is a TSA program that employs Behavioral Detection Officers (BDO) at airport terminals for the purpose of detecting behavioral based indicators of threats to aviation security.

The hearing will examine the state of behavioral science as it relates to the detection of terrorist threats to the air transportation system, as well as its utility to identify criminal offenses more broadly. The hearing will examine several independent reports-one by the Government Accountability Office (GAO), two by the National Research Council, and a number of Defense and Intelligence Community advisory board reports on the state of behavioral science relative to the detection of emotion, deceit, and intent in controlled laboratory settings, as well as in an operational environment. The Subcommittee will evaluate the initial development of the SPOT program, the steps taken to validate the science that form the foundation of the program, as well as the capabilities and limitations of using behavioral science in a transportation setting. More broadly, the hearing will also explore the behavioral science research efforts throughout DHS.

Background

The terrorist attacks on September 11, 2001 exposed a vulnerability in the nation's air transportation system. In order to augment other screening processes and procedures, TSA conducted operational testing of behavior detection techniques at a limited number of airports in October 2003.[1] In 2007, TSA created new BDO positions as part of the SPOT program with the goal of identifying persons who may pose a potential security risk by using behavioral indicators such as stress, fear, or deception.[2]

The indicators BDOs use form a checklist with corresponding values and thresholds. These indicators, values, and thresholds are used to assess passengers while in line awaiting security screening. When an individual displays behaviors or an appearance that exceeds a predetermined threshold, they are referred for additional screening. If, during the course of this secondary screening, individuals display behaviors that exceed another threshold, they are referred to law enforcement officers for further investigation.

Initially established to detect terrorist threats to the aviation transportation system,[3] the program's mission has since broadened to include the identification of behaviors indicative of criminal activity.[4] Critics of the program have argued that this expansion reflects the failure of the program to identify any terrorists, and therefore program success could only be quantified by broadening the goals to include crimi-

[1] Aviation Security: Efforts to validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges, Government Accountability Office, May 2010. Available at http://www.gao.gov/new.items/d10763.pdf

[2] *Ibid.*

[3] *Ibid.*

[4] Congressional Budget Justification FY2012, Department of Homeland Security.

nal activity which has a higher rate of occurrence.[5] This may or may not be a fair critique based on the extremely small sample size that terrorists would represent. Regardless of the rationale for the program's expanded scope, questions remain about whether indicators for terrorism are the same for criminal behavior.

As of March 2010, TSA employed roughly 3,000 BDOs at approximately 161 airports at a cost of $212 million a year.[6] In the President's fiscal year 2012 budget request, the Department seeks to add 175 more BDOs with an increase of $21 million - a 9.5 % increase over current funding levels.[7] In total, the five year budget profile for the SPOT program accounts for roughly $1.2 billion.[8]

Relevant Reviews

U.S. Government Accountability Office (GAO)

Aviation Security: Efforts to validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges

In May 2010, GAO issued a report titled "Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges" in response to a Congressional request to review the SPOT program. In preparing the report, GAO analyzed "(1) the extent to which TSA validated the SPOT program before deployment, (2) implementation challenges, and (3) the extent to which TSA measures SPOT's effect on aviation security."[9]

GAO issued the following findings associated with its review:

> *Although the Department of Homeland Security (DHS) is in the process of validating some aspects of the SPOT program, TSA deployed SPOT nationwide without first validating the scientific basis for identifying suspicious passengers in an airport environment. A scientific consensus does not exist on whether behavior detection principles can be reliably used for counterterrorism purposes, according to the National Research Council of the National Academy of Sciences. According to TSA, no other large-scale security screening program based on behavioral indicators has ever been rigorously scientifically validated. DHS plans to review aspects of SPOT, such as whether the program is more effective at identifying threats than random screening. Nonetheless, DHS's current plan to assess SPOT is not designed to fully validate whether behavior detection can be used to reliably identify individuals in an airport environment who pose a security risk. For example, factors such as the length of time BDOs can observe passengers without becoming fatigued are not part of the plan and could provide additional information on the extent to which SPOT can be effectively implemented. Prior GAO work has found that independent expert review panels can provide comprehensive, objective reviews of complex issues. Use of such a panel to review DHS's methodology could help ensure a rigorous, scientific validation of SPOT, helping provide more assurance that SPOT is fulfilling its mission to strengthen aviation security.*[10]

Additionally, GAO found issues relating to performance metrics, data integrity, and reach-back capabilities as well.

> *TSA is experiencing implementation challenges, including not fully utilizing the resources it has available to systematically collect and analyze the information obtained by BDOs on passengers who may pose a threat to the aviation system. TSA's Transportation System Operations Center has the resources to investigate aviation threats but generally does not check all law enforcement and intelligence databases available to it to identify persons referred by BDOs. Utilizing existing resources would enhance TSA's ability to quickly verify passenger identity and could help TSA to more reliably "connect the dots." Further, most BDOs lack a mechanism to input data on suspicious passengers into a database used by TSA analysts and also lack a means to obtain information from the Transportation System Operations Center on a timely basis. TSA states that it is in the process of providing input capabilities, but does not have a time frame*

---

[5] Weinberger, Sharon, "Intent to Deceive" Can the Science of Deception Detection Help to Catch Terrorists?" *Nature*, Vol. 465127, May 26, 2010, available at: http://www.nature.com/news/2010/100526/pdf/465412a.pdf

[6] *Supra* n.1.

[7] *Supra* n.4.

[8] *Supra* n.1.

[9] *Ibid.*

[10] *Ibid.*

for when this will occur at all SPOT airports. Providing BDOs, or other TSA personnel, with these capabilities could help TSA "connect the dots" to identify potential threats.

Although TSA has some performance measures related to SPOT, it lacks outcome-oriented measures to evaluate the program's progress toward reaching its goals. Establishing a plan to develop these measures could better position TSA to determine if SPOT is contributing to TSA's strategic goals for aviation security. TSA is planning to enhance its evaluation capabilities in 2010 to more readily assess the program's effectiveness by conducting statistical analysis of data related to SPOT referrals to law enforcement and associated arrests. [11]

Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue

In March of 2011, GAO issued a report to Congress in response to a new statutory requirement that GAO identify federal programs, agencies, offices, and initiatives, either within departments or governmentwide, which have duplicative goals or activities. The report contained a section on SPOT and stated:

Congress may wish to consider limiting program funding pending receipt of an independent assessment of TSA's SPOT program. GAO identified potential budget savings of about $20 million per year if funding were frozen at current levels until validation efforts are complete. Specifically, in the near term, Congress could consider freezing appropriation levels for the SPOT program at the 2010 level until the validation effort is completed. Assuming that TSA is planning to expand the program at a similar rate each year, this action could result in possible savings of about $20 million per year, since TSA is seeking about a $20 million increase for SPOT in fiscal year 2011. Upon completion of the validation effort, Congress may also wish to consider the study's results-including the program's effectiveness in using behavior-based screening techniques to detect terrorists in the aviation environment-in making future funding decisions regarding the program. [12]

Credibility Assessment at Portals Report

In April 2009, the Portals Committee issued a report for the Defense Academy for Credibility Assessment titled: "Credibility Assessment at Portals." [13] The committee recognized the need for "advanced and accurate credibility assessment," [14] which is described as "a decision making process whereby a communication is assessed as to its veracity." The Portals Committee had the following to say about SPOT:

"The adoption of SPOT occurred despite the fact that no study in the peer-reviewed scientific literature suggests that accurate credibility assessments can be made from unstructured observations. Within SPOT it appears that the observers are attempting to assess airline passengers by casual observation of facial micro-expressions (Wilber & Nakashima, 2007). There are several problems with this. First, scientific research does not support the notion that microexpressions reliably betray concealed emotion (Porter & ten Brinke, 2008). Second, whereas brief facial activity may reveal the purposeful manipulation of a felt emotion (Porter & ten Brinke, 2008), the problems of interpretation of such manipulation renders the approach useless for practical purposes. Third, the microexpression approach equates deception with manipulated emotion. This conceptual confusion obscures the fact that most forensically relevant lies are not lies about feelings but about actions in the past, present or future. In conclusion, the use of microexpressions to establish credibility is theoretically flawed and has not been supported by sound scientific research (Vrij, 2008)." [15]

JASON

Comprised of world renowned scientists, JASON advises the federal government on science and technology issues. The vast majority of its work is done at the re-

---

[11] *Ibid.*

[12] Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue, Government Accountability Office, March 2011, available at: http://www.gao.gov/new.items/d11318sp.pdf

[13] "Credibility Assessment at Portals," Portals Committee Report, April 17, 2009, available at: http://truth.boisestate.edu/eyesonly/Portals/PortalsCommitteeReport.pdf

[14] *Ibid.*

[15] *Ibid.*

quest of the Department of Defense and the intelligence community, so its reports are typically classified.

However, a 2010 Nature article that discusses the SPOT program in a piece on deception detection provides the following: "No scientific evidence exists to support the detection or inference of future behaviour, including intent,' declares a 2008 report prepared by the JASON defense advisory group."[16]

National Research Council (NRC) of the National Academies

Workshop Summary on *Field Evaluation in the Intelligence and Counterintelligence Context*

On September 22-23, 2009, the NRC's Board on Behavioral, Cognitive, and Sensory Sciences held a workshop on "the field evaluation of behavioral and cognitive sciences-based methods and tools for use in the areas of intelligence and counter intelligence."[17] The workshop was sponsored by the Defense Intelligence Agency and the Office of the Director of National Intelligence. The purpose of the workshop was to "discuss the best ways to take methods and tools from behavioral science and apply them to work in intelligence operations. More specifically, 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."[18]

The NRC published a report in 2010 summarizing the presentations and discussions over the 2-day period. Participants of the workshop included NRC members and experts in the behavioral sciences and intelligence community. The goal of the workshop was "not to provide specific recommendations but to offer some insight - in large part through specific examples taken from other fields - into the sorts of issues that surround the area of field evaluations. The discussions covered such ground as the obstacles to field evaluation of behavioral science tools and methods, the importance of field evaluation, and various lessons learned from experience with field evaluation in other areas."[19]

While the report identified several obstacles, one of interest to this Subcommittee hearing is "the pressure to use new devices and techniques as soon as they become available, without waiting for rigorous validation. Because lives are at stake, those in the field often push to adopt new methods and tools as quickly as possible and before there has been time to evaluate them adequately. Once a method is in widespread use, anecdotal evidence can lead its users to believe in its effectiveness and to resist rigorous testing, which may show that it's not as effective as they think."[20]

Protecting Individual Privacy in the Struggle Against Terrorists - A Framework for Program Assessment

From 2005 to 2007, the NRC's 21-member Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals held several meetings to "examine the role of data mining and behavioral surveillance technologies in counterterrorism programs."[21] The ensuing NRC report provides "a framework for making decisions about deploying and evaluating those [programs] and other information based programs on the basis of their effectiveness and associated risks to personal privacy."[22]

The report presented 13 conclusions and 2 broad recommendations. Of interest to this Subcommittee hearing are the following conclusions:

- *"Conclusion 3: Inferences about intent and/or state of mind implicate privacy issues to a much greater degree than do assessments or determinations of capability.*

  Although it is true that capability and intent are both needed to pose a real threat, determining intent on the basis of external indicators is inherently a much more subjective enterprise than determining capability. Determining intent or

---

[16] *Supra n.5.*

[17] "Field Evaluation in the Intelligence and Counterintelligence Context," National Research Council of the National Academies , 2010, available at: http://books.nap.edu/openbook.php?record__id=12854&page=R1

[18] *Ibid.*

[19] *Ibid.*

[20] ?Field Evaluation in the Intelligence and Counterintelligence Context,? National Research Council of the National Academies, March 2010, available at: http://www7.nationalacademies.org/bbcss/Highlights-Field%20Evaluation%20in%20the%20Intelligence%20and%20Counterintelligence%20Context.pdf

[21] "Protecting Individual Privacy in the Struggle against Terrorists - A Framework for Program Assessment," National Research Council of the National Academies, 2008, available at: http://books.nap.edu/openbook.php?record__id=12452&page=1

[22] *Ibid.*

state of mind is inherently an inferential process, usually based on indicators such as whom one talks to, what organizations one belongs to or supports, or what one reads or searches for online. Assessing capability is based on such indicators as purchase or other acquisition of suspect items, training, and so on. Recognizing that the distinction between capability and intent is sometimes unclear, it is nevertheless true that placing people under suspicion because of their associations and intellectual explorations is a step toward abhorrent government behavior, such as guilt by association and thought crime. This does not mean that government authorities should be categorically proscribed from examining indicators of intent under all circumstances-only that special precautions should be taken when such examination is deemed necessary."

- *"Conclusion 4: Program deployment and use must be based on criteria more demanding than 'it's better than doing nothing."*

In the aftermath of a disaster or terrorist incident, policy makers come under intense political pressure to respond with measures intended to prevent the event from occurring again. The policy impulse to do something (by which is usually meant something new) under these circumstances is understandable, but it is simply not true that doing something new is always better than doing nothing. Indeed, policy makers may deploy new information-based programs hastily, without a full consideration of (a) the actual usefulness of the program in distinguishing people or characteristic patterns of interest for follow-up from those not of interest, (b) an assessment of the potential privacy impacts resulting from the use of the program, (c) the procedures and processes of the organization that will use the program, and (d) countermeasures that terrorists might use to foil the program.

- *"Conclusion 10: Behavioral and physiological monitoring techniques might be able to play an important role in counterterrorism efforts when used to detect (a) anomalous states (individuals whose behavior and physiological states deviate from norms for a particular situation) and (b) patterns of activity with well-established links to underlying psychological states.*

Scientific support for linkages between behavioral and physiological markers and mental state is strongest for elementary states (simple emotions, attentional processes, states of arousal, and cognitive processes), weak for more complex states (deception), and nonexistent for highly complex states (terrorist intent and beliefs). The status of the scientific evidence, the risk of false positives, and vulnerability to countermeasures argue for behavioral observation and physiological monitoring to be used at most as a preliminary screening method for identifying individuals who merit additional follow-up investigation. Indeed, there is no consensus in the relevant scientific community nor on the committee regarding whether any behavioral surveillance or physiological monitoring techniques are ready for use at all in the counterterrorist context given the present state of the science."

- *"Conclusion 11: Further research is warranted for the laboratory development and refinement of methods for automated, remote, and rapid assessment of behavioral and physiological states that are anomalous for particular situations and for those that have well-established links to psychological states relevant to terrorist intent.*

A number of techniques have been proposed for the machine-assisted detection of certain behavioral and physiological states. For example, advances in magnetic resonance imaging (MRI), electroencephalography (EEG), and other modern techniques have enabled measures of changes in brain activity associated with thoughts, feelings, and behaviors. Research in image analysis has yielded improvements in machine recognition of faces under a variety of circumstances (e.g., when a face is smiling or when it is frowning) and environments (e.g., in some nonlaboratory settings).

However, most of the work is still in the basic research stage, with much of the underlying science still to be validated or determined. If real-world utility of these techniques is to be realized, a number of issues- practical, technical, and fundamental-will have to be addressed, such as the limits to understanding, the largely unknown measurement validity of new technologies, the lack of standardization in the field, and the vulnerability to countermeasures. Public acceptability regarding the privacy implications of such techniques also remains to be demonstrated, especially if the resulting data are stored for unknown future uses or undefined lengths of time.

For example, the current state-of-the-art of functional MRI technology can identify changes in the hemodynamics in certain regions of the brain, thus signaling activ-

ity in those regions. But such results are not necessarily consistent across individuals (i.e., different areas in the brains of different individuals may be active under the same stimulus) or even in the same individual (i.e., a slightly different part of the brain may become active even in the same individual under the same stimulus). Certain regions of the brain may be active under a variety of different stimuli.

In short, understanding of what these regions do is still primitive. Furthermore, even if simple associations can be made reliably in laboratory settings, this does not necessarily translate into usable technology in less controlled situations. Behavior of interest to detect, such as terrorist intent, occurs in an environment that is very different from the highly controlled behavioral science laboratory."

- *"Conclusion 12: Technologies and techniques for behavioral observation have enormous potential for violating the reasonable expectations of privacy of individuals.*

Because the inferential chain from behavioral observation to possible adverse judgment is both probabilistic and long, behavioral observation has enormous potential for violating the reasonable expectations of privacy of individuals. It would not be unreasonable to suppose that most individuals would be far less bothered and concerned by searches aimed at finding tangible objects that might be weapons or by queries aimed at authenticating their identity than by technologies and techniques whose use will inevitably force targeted individuals to explain and justify their mental and emotional states. Even if behavioral observation and physiological monitoring are used only as a preliminary screening methods for identifying individuals who merit additional follow-up investigation, Because the inferential chain from behavioral observation to possible adverse judgment is both probabilistic and long, behavioral observation has enormous potential for violating the reasonable expectations of privacy of individuals. It would not be unreasonable to suppose that most individuals would be far less bothered and concerned by searches aimed at finding tangible objects that might be weapons or by queries aimed at authenticating their identity than by technologies and techniques whose use will inevitably force targeted individuals to explain and justify their mental and emotional states. Even if behavioral observation and physiological monitoring are used only as a preliminary screening methods for identifying individuals who merit additional follow-up investigation, these individuals will be subject to suspicion that would not fall on others not so identified."[23]

Issues

Detection of Emotion

The state of science relative to the detection of emotion, deceit, and intent are vastly different. Decades of research have been devoted to the detection of emotion using verbal, nonverbal, and microfacial expressions. Each of these observational techniques have shown to have varying degrees of success at determining an individual's emotion, but generally speaking, a scientific foundation does exist to support the assertion that emotion can be determined through behavioral cues.

Detection of Deceit

The foundation of research for detecting an expression of deceit is rooted in that of emotion. For example, it is posited that a deceitful person would express emotions such as stress, and that stress can be attributed to concealing a lie. The state of the science in this regard is less solid. Witnesses at the hearing will testify to the current strengths and weaknesses of this field.

Detection of Intent

Even less certainty exists regarding the ability to determine intent. This ability is asserted by assuming that a person who intends to do harm will be concealing this fact, thereby expressing deceitful behaviors - and that deceitful behavioral cues are founded in stress, which in turn are displayed in emotion. This chain of reasoning takes the underlying assumption that behavioral indicators exist for detecting emotion and infers that indicators can therefore be used to detect deceit, and therefore intent. Very little, if any, evidence exists in the scientific literature to support this hypothesis, yet this is the goal of the SPOT program - to identify individuals who may pose a threat to aviation security.

---

[23] *Ibid.*

Laboratory vs. Operational Settings

The vast preponderance of behavioral science research conducted relative to the detection of emotion, deceit, and intent has been done in a laboratory setting. As the National Research Council noted in its 2008 report, "Behavior of interest to detect, such as terrorist intent, occurs in an environment that is very different from the highly controlled behavioral science laboratory."[24]

Utility for Counterterrorism

Even if one was to stipulate that a body of evidence existed to support the claim that one could detect intent using behavioral indicators, it remains to be seen how useful this would be in a counterterrorism context. In all likelihood, anyone seeking to cause harm would employ countermeasures designed to conceal their emotions. It remains to be seen what impact countermeasures will have on the ability to detect emotions, deception, or intent, but if other deception detection tools (such as the polygraph) are any indicator, they could severely degrade the capability.

Utility in a U.S. Aviation Transportation Setting

The SPOT program is loosely based on the Israeli model successfully employed by El Al Airlines. This highly successful program employs more agents in more locations throughout the airport, conducts multiple face to face interviews, actively profiles passengers, and operates in smaller and fewer airports. They also have much fewer passengers and far fewer flights than the U.S. air transportation system. Israeli screeners also receive more training than the four days of classroom training, and three days of on the job training that BDOs receive. Scaling up such an enterprise to accommodate the U.S. Aviation Transportation Sector would severely restrict the flow of commerce and passengers.

DHS S&T Validation

In its report, GAO states that "TSA deployed SPOT nationwide without first validating the scientific basis for the program."[25] To its credit, DHS S&T initiated a review two and a half years ago to "determine whether SPOT is more effective at identifying passengers who may be threats to the aviation system than random screening."[26] GAO goes on to point out in its report, "However, S&T's current research plan is not designed to fully validate whether behavior detection and appearances can be effectively used to reliably identify individuals in an airport terminal environment who pose a risk to the aviation system."[27] The report further states that, according to the National Research Council, "an independent panel could provide an objective assessment of the methodologies and findings of DHS's study to better ensure that SPOT is based on valid science."[28]

These are two important points. First, the S&T review is not designed to validate the underlying behavioral cues, but rather to simply demonstrate whether the program, as a whole, is more successful than random sampling. As GAO stated in its recent "Duplication" report, "DHS's response to GAO's report did not describe how the review currently planned is designed to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program."[29] Second, based on the Statement of Work associated with S&T's review, questions remain as to whether or not the review is truly independent.

The Statement of Work affirms that S&T had a direct role in selecting peer reviewers, as well as planning and structuring workshops that informed the methodology to validate the program. The Statement of Work also afforded DHS the ability to review and provide revision recommendations at numerous points in the process. Finally, the Statement of Work indicates that deliverables are to be provided to S&T directly.[30] Whether or not this affected the outcome is uncertain. The validation work was conducted by the American Institute for Research, a high respected and reputable firm, but ultimately they are contractually bound by the parameters and scope defined by Statement of Work negotiated with DHS. It remains to be seen whether the review was an independent assessment, as recommended by the National Research Council, or more of a collaboration.

---

[24] *Supra* n.21.
[25] *Supra* n.1.
[26] *Ibid.*
[27] *Ibid.*
[28] *Ibid.*
[29] *Supra* n.12.
[30] Statement of Work for the Naval Research Laboratory, Project Hostile Intent: Behavioral-Based Screening Indicators Validation, U.S. department of Homeland Security, Science and Technology Directorate, Human Factors and Behavioral Sciences Division, PR# RSHF-11-00007.

Nevertheless, S&T's two and a half year review (at a cost of $2.5 million) was initially planned to be delivered in Fiscal year 2011,[31] then February 2011,[32] and then the end of March 2011. Its current release date is for April 8th, two days after our hearing. The Subcommittee postponed this hearing, initially scheduled for March 17th, for a number of reasons, including allowing S&T more time to produce the report.

Witnesses

- **Mr. Stephen Lord**, Director, Homeland Security and Justice Issues, Government Accountability Office
- **Transportation Security Administration** (Invited)
- **Mr. Larry Willis**, Program Manager, Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security
- **Dr. Paul Ekman**, Professor Emeritus of Psychology, University of California, San Francisco, and President and Founder, Paul Ekman Group, LLC
- **Dr. Maria Hartwig**, Associate Professor, Department of Psychology, John Jay College of Criminal Justice
- **Dr. Philip Rubin**, Chief Executive Officer, Haskins Laboratories
- **Lieutenant Detective Peter J. DiDomenica**, Boston University Police

[31] *Supra* n.1.
[32] *Supra* n.12.

Appendix 1

Department of Homeland Security
Science and Technology Directorate
Human Factors Behavioral Sciences Projects

These projects advance national security by developing and applying the social, behavioral, and physical sciences to improve identification and analysis of threats, to enhance societal resilience, and to integrate human capabilities into the development of technology.

Commercial Data Sources Project

Project Manager: Patty Wolfhope
Project Overview: The Science and Technology (S&T) Directorate Human Factors Behavior Sciences Division (HFD) Commercial Data Sources Project will quantitatively assess the utility of commercial data sources to augment governmentally available information about people, foreign and domestic, being screened, investigated, or vetted by the Department. The use of commercial data sources may provide a valuable source of corroborating information to ensure that an individual's identity and eligibility for a particular license, privilege, or status is correctly evaluated during screening. This project is part of the Personal Identification Systems Thrust Area and Credentialing Program within HFD.

Community Perceptions of Technology Panel Project

Project Manager: Ji Sun Lee

Project Overview: The Science and Technology (S&T) Directorate Human Factors/Behavioral Sciences Division (HFD) Community Perceptions of Technology Panel (CPT) Project brings together representatives of industry, public interest, and community-oriented organizations to better understand and integrate community perspectives and concerns in the development, deployment, and public acceptance of technology. This will yield feedback to aid ongoing technology and process development and strategies to accurately inform the public of new approaches to securing the homeland. This is designed to better ensure acceptance of the technology within affected communities. This project is part of the Human Technology Integration Thrust Area and Technology Acceptance and Integration Program within HFD.

Community Resilience Project

Project Manager: Michael Dunaway

Project Overview: The Science and Technology (S&T) Directorate Human Factors/Behavioral Sciences Division (HFD) Counter-Improvised Explosives Devices (IED) Community Resilience Project conducts research into methodologies for effective hazard and risk communications to enhance the ability of local officials to convey understandable and credible warnings of IED activity to the public. This project will help local government and civic officials understand how to properly frame risk warnings and post-event instructions to the public in a manner that maximizes the public's understanding of the instructions provided and maintains public trust and confidence. HFD is executing this project as part of the Counter Improvised Explosive Devices (C-IED) Thrust Area and Mitigate Program within Explosives Division.

Counter-IED Actionable Indicators and Countermeasures Project

Project Manager: Allison Smith, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors/Behavioral Sciences Division (HFD) Counter-Improvised Explosives Devices (IED) Actionable Indicators and Countermeasures Project supports the intelligence and law enforcement communities in identifying actors that pose significant IED threats in the United States homeland. This project will provide practical tools through the synthesis of state-of-the-art social and behavioral science databases, case studies, surveys, and fieldwork and advanced computational modeling, simulation, and visualization technologies. It will also provide policymakers with scientifically tested strategies to prevent radicalization and IED attacks before they occur by examining how social and behavioral science principles can support the development of counter-radicalization efforts. HFD is executing this project as part of the Counter Improvised Explosive Devices (C-IED) Thrust Area and Prevent/Deter Program.

Credentialing Project

Project Manager: Patty Wolfhope

Project Overview: The Science and Technology (S&T) Directorate Human Factors Behavior Sciences (HFD) Division Credentialing Project develops tamper-proof credentialing systems that incorporate biometric information; such as a biometrics-based card-and-reader system. The project developed a laboratory test and evaluation protocol for the transportation worker identification card (TWIC) reader and plans to initiate research and design activities to improve the range and reliability of secure contactless technologies. This project is part of the Personal Identification Systems Thrust Area and Credentialing Program within HFD.

Enhanced Screener – Technology Interface Project

Project Manager: Josh Rubinstein, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors Behavioral Sciences (HFD) Division Enhanced Screener-Technology Interface Project characterizes screener-performance issues, proposes new screener technologies and procedures, and develops training curricula to optimize security effectiveness and reduce human fatigue and injury, while reducing training requirements and overall cost. This project is part of the Human Technology Integration Thrust Area and Transportation Technology-Human Integration Program within HFD.

Enhancing Public Response and Community Resilience Project

Project Manager: Michael Dunaway

Project Overview: The Science and Technology (S&T) Directorate Human Factors/Behavioral Sciences Division (HFD) Enhancing Public Response and Community Resilience Project examines public needs (shelter, food, disaster relief, etc.) that arose during the evacuation from southern Texas during Hurricanes Katrina and Rita in order to enhance federal, state, local and private sector response to future catastrophic events. The goal is to capture and communicate lessons learned to enhance federal, state, local and private sector responses to future catastrophic events. This project is part of the Social and Behavioral Threat Analysis (SBTA) Thrust Area and Community Preparedness and Resilience Program within HFD.

High Impact Technological Solution – Biometric Detector Project

Project Manager: Arun Vemury

Project Overview: The Science and Technology (S&T) Directorate High Impact Technological Solutions (HITS) Project executed by the Human Factors/Behavioral Science Division (HFD) will provide efficient, high quality, contact less acquisition of fingerprint biometric signatures for identity management. This will result in significantly improved throughput and signal quality, thereby improving recognition and reducing false positive rates. The goal is to develop a fingerprint acquisition device that can be transitioned for implementation across Department components. This project is part of the Innovations Portfolio/Homeland Security Advanced Research Project Agency Program (HSARPA) within the S&T Directorate.

Homeland Innovation Prototypical Solutions – Future Attribute Screening Technology (FAST) Project

Project Manager: Bob Burns

Project Overview: The Homeland Security Advanced Research Project Agency (HSARPA) and Science and Technology (S&T) Directorate Human Factors/Behavioral Sciences Division (HFD) Future Attribute Screening Technology (FAST) Project is an initiative to develop innovative, non-invasive technologies to screen people at security checkpoints. FAST is grounded in research on human behavior and psychophysiology, focusing on new advances in behavioral/human-centered screening techniques. The aim is a prototypical mobile suite (FAST M2) that would be used to increase the accuracy and validity of identifying persons with malintent (the intent or desire to cause harm). Identified individuals would then be directed to secondary screening, which would be conducted by authorized personnel. This project is part of the Innovations Portfolio/Homeland Security Advanced Research Project Agency (HSARPA) Program within the S&T Directorate.

Hostile Intent Detection – Automated Prototype Project

Project Manager: Larry Willis

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Hostile Intent Detection - Automated Prototype Project demonstrates real-time automated intent detection using non-invasive and culturally neutral behavioral indicators. S&T plans to transition the automated hostile intent prototype to the Transportation Security Administration, Customs and Border Protection, and Immigration and Customs Enforcement. This project is a part of the Social and Behavioral Threat Analysis Thrust Area and Suspicious Behavior Detection Program within HFD.

Hostile Intent Detection – Training & Simulation Project

Project Manager: Larry Willis

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Hostile Intent Detection - Training and Simulation Project develops computer-based simulation to train behavior-based stand-off detection for future hostile intent using indicators from the interactive screening environment (Hostile Intent Detection - Automated Prototype) and the observational environment (Hostile Intent Detection - Validation) to support screening and interviewing interactions at air, land, and maritime portals. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Suspicious Behavior Detection Program within HFD.

Hostile Intent Detection – Validation Project

Project Manager: Larry Willis

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Hostile Intent Detection - Validation Project provides cross-cultural validation of behavioral indicators employed by Department of Homeland Security's operational components to screen passengers at air, land, and maritime ports. The project will integrate these validated behavioral indicators into the screening curriculum of each component's existing training program. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Suspicious Behavior Detection Program within HFD.

Human Systems Engineering Project

Project Managers: Darren P. Wilson and Janae Lockett-Reynolds, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Project develops, demonstrates and evaluates a standardized process for implementing human systems integration. It will focus on defining human performance requirements in the development of systems and technology, and on methods and measures needed to evaluate existing technology in terms of human performance requirements. This effort also will result in greater understanding of the needs of the various Department end-user communities, as well as developing tools to best identify how to recruit, select, train, support, and retain operational staff. A systematic approach based on the integration of the human component will lead to enhanced system design, safety, efficiency, and operational performance. This project is part of the Human Technology Integration Thrust Area and Human Systems Research and Engineering Program within HFD.

Human Systems Engineering Research Project

Project Manager: Jennifer O'Connor, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Science Division (HFD) projects examine human perception and ability to detect targets and threats as they pertain to the design of systems that maximize human performance, and the effectiveness of the technology operators use in the field. Results of this research allow the program to focus more closely on the psychological determiners that impact successful discrimination of threats and reduce false alarms. In addition to focusing on human perception, the project will also address how humans process information and how that impacts the human-machine interface. This project is part of the Human Technology Integration Thrust Area and Human Systems and Engineering Program within HFD.

Insider Threat Detection Program

Project Manager: Jennifer O'Connor, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Insider Threat Detection Project will detect in-

sider behavior that is likely to present or lead to a threat to critical infrastructure using behavioral indicators. Department of Homeland Security will collaborate with other U.S. agencies and international partners to move beyond the current focus on responses to accomplished hostile insider acts, and begin developing a greater capacity to deter and detect insider threats before substantial harm has been done. The immediate operational goal is to produce new and better tools to identify behavior patterns and characteristics identifiable before, during, and after employment that are associated with insider threats. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Suspicious Behavior Detection Program withinHFD.

Mobile Biometrics System Project

Project Manager: Patty Wolfhope

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavior Sciences Division (HFD) Mobile Biometrics Project develops prototype technologies for mobile biometrics screening at remote sites along U.S. borders, during disasters and terrorist incidents, at sea, and in other places where communications access is limited. The goal is to demonstrate mobile biometrics screening capabilities and technologies that meet the future needs of Department operational users, but currently are not available with conventional biometrics systems. This project is part of the Personal Identification Systems Thrust Area and Biometrics Program within HFD.

Multi-modal Biometrics Project

Project Manager: Arun Vemury

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavior Sciences Division (HFD) Multi-modal Biometrics Project develops biometric technologies that accurately and rapidly identify individuals. The operational goal is to provide the capability to non-intrusively collect two or more biometrics (fingerprint, face image, and iris recognition) in less than ten seconds at a ninety-five percent acquisition rate without impeding the movement of individuals. The multi-modal technology will allow the Department to compare and match biometric samples from different sources, collected with different sensor technologies, under varying environmental conditions -- a capability that eludes existing technology. This project is part of the Personal Identification Systems Thrust Area and Biometrics Program within HFD.

Muslim Community Integration Project

Project Manager: Allison Smith, Ph.D.

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Muslim Community Integration Project conducts ethnographic research to examine the experiences of Muslims and non-Muslims in several communities throughout the U.S. The project will provide insights into the current state of Muslim communities focusing on their role and status in America and their perceptions of American society. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Community Preparedness, Response and Recovery Program within HFD.

Predictive Screening Project

Project Manager: Larry Willis

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Counter-Improvised Explosives Devices (Counter-IED) Predictive Screening Project will derive observable behaviors that precede a suicide bombing attack and develop extraction algorithms to identify and alert personnel to indicators of suicide bombing behavior. HFD is executing this project as part of the Counter-IED Thrust Area and Predict Program.

Risk Prediction Project

Project Manager: Larry Willis

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Counter-Improvised Explosives Devices Risk Prediction Project will develop high speed software to identify improvised explosive device (IED) target and staging areas based upon group-and-cultural-specific tactics, techniques, and procedures derived from past foreign attacks. The goal is to use this

information to prioritize the risk of likely potential targets of IED attacks within the United States. HFD is executing this project as part of the Counter-IED Thrust Area and Predict Program.

Social Network Analysis for Community Resilence Project

Project Manager: Michael Dunaway

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Social Network Analysis for Community Resilience Project develops a modeling capability for identifying formal and informal social networks that may be useful in enhancing preparedness and community resilience to natural disasters and terrorist events. This effort will leverage social network analysis research for understanding terrorist networks, social and financial transactions, and the spread of infectious diseases, and apply that knowledge to the construction of networks dedicated to strengthening local response capabilities and preparedness. It will also leverage past and on-going work from the Department of Defense (DOD) and other agencies. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Community Preparedness and Resilience Program within HFD.

Violent Intent Modeling and Simulation Project

Project Manager: Ji Sun Lee

Project Overview: The Science and Technology (S&T) Directorate Human Factors/ Behavioral Sciences Division (HFD) Violent Intent Modeling and Simulation Project develops intelligence analysis frameworks, including extraction of terrorist intention signatures, systematic estimation of future terrorist behavior based on social and behavioral sciences, and modeling and simulations of future terrorist behavior influences. It identifies leading edge social science modeling and simulation technologies and advances social science modeling and data fusion capabilities in such areas as hybrids of neural nets, structural equations, genetic algorithms, social networks, etc. This project is part of the Social and Behavioral Threat Analysis Thrust Area and Motivation and Intent Program within HFD.

Source: http://www.dhs.gov/files/programs/gc__1218480185439.shtm

Chairman BROUN. The Subcommittee on Investigations and Oversight will come to order. Good morning. Welcome to today's hearing titled "Behavioral Science and Security: Evaluating TSA's SPOT Program." You will find in front of you packets containing our witness panel's written testimony, biographies, and Truth-in-Testimony disclosures.

Before we get started, this being the first meeting of the Investigations and Oversight Subcommittee for the 112th Congress, I would like to ask the Subcommittee's indulgence to introduce myself. It is an honor and a pleasure for me to chair the I&O Subcommittee for this Congress, and it is a position that I do not take lightly. I want all Members of this Subcommittee to know that my door is always open, that I will endeavor to serve all Members fairly and impartially, and that I will work to serve the best interests of Congress, and all Americans, to ensure that the agencies and programs under our jurisdiction are worthy of the public's support.

And I recognize myself for five minutes for an opening statement. Today the Subcommittee meets to evaluate TSA's SPOT program. Developed in the wake of September 11, 2001, it was deployed on a limited basis in a select number of airports in 2003. In 2007, TSA created new Behavioral Detection Officer (BDO) positions whose goal was to use behavioral indicators to identify persons who may pose a potential security risk to aviation. This goal expanded in recent years to include the identification of any criminal activity. TSA currently employs about 3,000 BDOs in about 161 airports at the cost of over $200 million a year. The President's fiscal year 2012 budget request asks for an increase of 9.5 percent and an additional 175 BDOs. Over the next five years, the SPOT program will cost roughly $1.2 billion.

Outside of a few brief exchanges at Appropriations Committee hearings, Congress has not evaluated this program. That isn't to say that Congress wasn't paying attention, as GAO conducted a comprehensive review that culminated in a report on the SPOT program last May. In that report, GAO identified several problems with the program, most notably that it was deployed without being scientifically validated.

This is a common theme that this Committee is increasingly forced to deal with. Expensive programs are rolled out without conducting the necessary analysis. This has become a trend throughout the Federal Government but particularly at the Department of Homeland Security.

This Committee has a long history with the development and acquisition of the Advanced Spectroscopic—as a southerner it is hard to say Spectroscopic—Portal program, but other technology programs such as the Backscatter Advanced Imaging Technology, explosives trace-detection portal machines, and the Cargo Advanced Automated Radiography System all ran into problems because they were rolled out before they were ready. DHS either fails to properly test and evaluate the technology, does not conduct a proper risk analysis, or neglects to conduct a cost/benefit analysis.

A crucial aspect that is oftentimes taken for granted by DHS is the nexus between those developing the technology and those actually using it. In the case of SPOT, it seems as though the operators got out ahead of the developers, but typically what we see is the

opposite; the scientists and engineers developing capabilities that do not appropriately fit into an operational environment. Unfortunately, this is an issue that the Committee is unable to address today because of TSA's refusal to attend.

The goal of this hearing is to shed light on the processes by which DHS created the SPOT program, to better understand the state of the science that forms the foundation of the program, to examine the methodologies by which DHS S&T is evaluating the program, and to identify any opportunities to improve how behavioral sciences are utilized in the security context. The goal is not to throw out the proverbial baby with the bath water, but rather to ensure that the science being used is not oversold or undersold.

SPOT is the first behavioral science program to stick its neck out for evaluation. This review is an opportunity to look at how behavioral sciences can be used appropriately across the security enterprise and to understand its limitations and strengths.

To its credit, DHS S&T is conducting an evaluation of the program for TSA. This report was due earlier this year in February, then at the end of March, and now is expected shortly. And hopefully we will get that shortly. While this is a good first step, I am eager to hear how independent this evaluation truly is. I look forward to understanding the review's methodology, its assumptions, and what level of input and access DHS S&T had in its design, formulation, and findings.

As GAO stated in its recent duplication report, "DHS's response to GAO's report did not describe how the review currently planned is designed to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program." I hope you all understood that bureaucratese.

The use of behavioral sciences in the security setting is not just another layer to security. There is clear opportunity costs that have to be paid. For every BDO employed to identify behaviors, there is one screener who is not looking at an x-ray of baggage, one intelligence analyst not employed, or one air marshal not in the sky. I realize this isn't a one-for-one substitute, but clearly there are tradeoffs that have to be made in a very difficult fiscal environment.

Also, I would be remiss if I did not address the clear privacy issues that this technology and other DHS technologies present. Privacy, along with the serious Constitutional questions I have, only compounds the complexity of the issue. While the focus of the hearing today is the science behind the program, I don't want these other important issues to be forgotten.

Now, the Chair recognizes Ms. Edwards for an opening statement. Ms. Edwards?

[The prepared statement of Mr. Broun follows:]

PREPARED STATEMENT OF CHAIRMAN PAUL BROUN

Today the Subcommittee meets to evaluate TSA's SPOT program. Developed in the wake of September 11, 2001, it was deployed on a limited basis in a select number of airports in 2003. In 2007, TSA created new Behavioral Detection Officer (BDO) positions whose goal was to use behavioral indicators to identify persons who may pose a potential security risk to aviation. This goal expanded in recent years to include the identification of any criminal activity. TSA currently employs about 3,000 BDOs in about 161 airports at a cost of over $200 million a year. The Presi-

dent's FY12 budget request asks for an increase of 9.5%, and an additional 175 BDOs. Over the next five years, the SPOT program will cost roughly $1.2 billion.

Outside of a few brief exchanges at Appropriations Committee Hearings, Congress has not evaluated this program. That isn't to say that Congress wasn't paying attention, as GAO conducted a comprehensive review that culminated in a report on the SPOT program last May. In that report, GAO identified several problems with the program, most notably that it was deployed without being scientifically validated.

This is a common theme that this Committee is increasingly forced to deal with. Expensive programs are rolled out without conducting the necessary analysis. This has become a trend throughout the federal government, but particularly at the Department of Homeland Security. This Committee has a long history with the development and acquisition of the Advanced Spectroscopic Portal program, but other technology programs such as the Backscatter Advanced Imaging Technology, explosives trace-detection portal machines, and the Cargo Advanced Automated Radiography System all ran into problems because they were rolled out before they were ready. DHS either fails to properly test and evaluate the technology, does not conduct a proper risk analysis, or neglects to conduct a cost-benefit analysis. A crucial aspect that is often times taken for granted by DHS is the nexus between those developing the technology, and those actually using it. In the case of SPOT, it seems as though the operators got out ahead of the developers, but typically what we see is the opposite, the scientists and engineers developing capabilities that do not appropriately fit into an operational environment. Unfortunately, this is an issue that the Committee is unable to address today because of TSA's refusal to attend.

The goal of this hearing is to shed light on the processes by which DHS created the SPOT program, to better understand the state of the science that forms the foundation of the program, to examine the methodologies by which DHS S&T is evaluating the program, and identify any opportunities to improve how behavioral sciences are utilized in the security context. The goal is not to "throw the baby out with the bath water," but rather to ensure that the science being used is not oversold, or undersold. SPOT is the first behavioral science program to stick its neck out for validation. This review is an opportunity to look at how behavioral sciences can be used appropriately across the security enterprise and to understand its limitations and strengths.

To its credit, DHS S&T is conducting an evaluation of the program for TSA. This report was due earlier this year in February, then at the end of March, and is now expected shortly. While this is a good first step, I am eager to hear how independent this evaluation truly is. I look forward to understanding the review's methodology, its assumptions, and what level of input and access DHS S&T had in its design, formulation and findings. As GAO stated in its recent duplication report, "DHS's response to GAO's report did not describe how the review currently planned is designed to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program."

The use of behavioral sciences in the security setting is not just another layer to security. There are clear opportunity costs that have to be paid. For every BDO employed to identify behaviors, there is one screener who is not looking at an x-ray of baggage, one intelligence analyst not employed, or one air marshal not in the sky. I realize this isn't a one-for-one substitute, but clearly there are trade-offs that have to be made in a very difficult fiscal environment. Also, I would be remiss if I did not address the clear privacy issues that this technology and other DHS technologies present. Privacy, along with the serious Constitutional questions I have, only compounds the complexity of the issue. While the focus of the hearing today is the science behind the program, I don't want these other important issues to be forgotten.

Ms. EDWARDS. Thank you, Mr. Chairman. And congratulations to you as you convene the first of what I hope are many oversight hearings to make sure that we are paying attention to the kind of oversight that we need to engage in on the Science and Technology Committee on behalf of the taxpayers.

I would like to say that I, too, am disappointed that TSA is not here today, wasn't able to provide a witness. I think they lost an important opportunity to inform the Congress and the public why they believe the SPOT program is worthy of our support. And I hope they will cooperate with this Committee and the Congress in the future. And I hope it is not terribly distracting as we get to the

witnesses. I don't want any one of them to be identified as TSA and I know it is a little confusing for me up here.

Let me just say in opening that I think each one of us has had an experience of instinctively sensing that something about a situation or person is wrong or it is worrying. Police officers, immigration officers, transportation security officers have those instinctive feelings all the time. However, it is an open question whether instinctive reactions are reliable as warnings of mal-intent. We also do not know whether a person can be trained to accurately sort through their instinctive reactions, choosing to intervene when faced with a potential threat and to resist reactions based on racial profiling.

What the Transportation Security Administration has tried to do is develop behavioral training for officers so they can quickly and accurately assess and screen passengers. Can hunches be harnessed in service of identifying potential threats to air safety? That is the key question that underlies today's hearing and I hope we will be able to dig deeply into those questions.

After Richard Reid's failed shoe bombing, some in the aviation security community concluded that we were spending too much time and money on trying to stop the bomb and not enough to stop the bomber. Screening of passengers by observation techniques, or SPOT, was viewed by TSA as a way to get some officers' eyes off the scanning screens and onto the passengers.

Those credited with helping to develop the SPOT program, some of whom are testifying before us today, intended the program to train Behavior Detection Officers (BDOs) to focus on an individual's behavior, appearance, and demeanor. An ongoing concern, however, with the BDOs and with law enforcement as well is that they not engage in racial profiling. If BDOs focus on a passenger's ethnic, religious, or racial qualities, they are violating the law, and they are not acting to protect the flying public.

Terrorists have come in all colors, shapes, and sizes, and if security personnel were fixated on a profiling approach to finding the next Mohammed Atta, then they would miss identifying the next John Walker Lindh, Timothy McVeigh, or Richard Reid.

The SPOT program tries to identify a specific menu of behaviors that will naturally emerge due to elevated levels of anxiety or stress. The hypothesis is that terrorists would display those cues when attempting to enter a secure facility such as an airport. But behavioral scientists do not agree on these nonverbal cues and they don't agree on whether terrorists would exhibit them. Because it is impossible to get a group of terrorists to participate in a double-blind experiment, it is hard to validate the theory.

DHS points to the program's success in identifying people who have violated the law and are caught, but no one can be certain criminals and terrorists behave in a similar fashion. TSA relies on nonverbal cues to help sort through the more than one million passengers that fly into the United States each day. Nonverbal cues provide a filtering method to allow officers to determine who they should engage in discussion looking for verbal signs of deception. There is more agreement among social scientists that verbal interactions with individuals can actually help in detecting deception.

We would hope that a DHS-funded validation report on the SPOT program would be available for this hearing today. That report purportedly shows that SPOT-trained Behavior Detection Officers are much more likely to identify what TSA deems as "high-risk passengers" as against a purely random sample of passengers. We look forward to the report's completion and its findings, but without it, we are missing an important initial assessment of the program's performance.

Over the past ten years since the 9/11 terrorist attacks, Congress has allocated billions of dollars to the Department of Homeland Security for the development of tools and technologies to keep our air travel secure. Too often that investment has been wasted and too often we have relied on technology that is not adequately tested before it is deployed. It is not based on adequate scientific evidence of effectiveness, and almost inevitably, the technology has proven costly to acquire, deploy, and service.

So I look forward to today's hearing and to asking questions about the more than $200 million a year that we are spending to make sure that we carefully evaluate SPOT's operational merit. And with that, I yield.

[The prepared statement of Ms. Edwards follows:]

PREPARED STATEMENT OF RANKING MEMBER DONNA F. EDWARDS

Every one of us has had the experience of instinctively sensing that something about a situation or a person is wrong, worrying. Police officers, immigration officers, Transportation Security Officers have those same instinctive feelings all the time. However, it is an open question whether instinctive reactions are reliable as warnings of mal-intent. We also do not know whether a person can be trained to accurately sort through their instinctive reactions, choosing to intervene when faced with a potential threat and to resist reactions based on racial profiling.

What the Transportation Security Administration (TSA) has tried to do is develop behavioral training for officers so that they can quickly and accurately screen passengers. Can hunches be harnessed in service of identifying potential threats to air traffic safety? That is the key question that underlies today's hearing.

After Richard Reid's failed shoe-bombing, some in the aviation security community concluded that we were spending too much time and money on trying to stop the bomb and not enough effort trying to stop the bomber. Screening of Passengers by Observation Techniques or SPOT was viewed by TSA as the way to get some officers' eyes off the scanning screens and onto the passengers.

Those credited with helping to develop the SPOT program, some of whom are testifying before us today, intended the program to train behavior detection officers (BDOs) to focus on an individual's behavior, appearance and demeanor. An ongoing concern with the BDOs, and with law enforcement as well, is that they not engage in racial profiling, If BDO's focus on a passenger's ethnic, religious or racial qualities they are violating the law, and they are not acting to protect the flying public. Terrorists have come in all colors, shapes and sizes. If security personnel were fixated on a profiling approach to finding the next Mohammed Alta, then they would miss identifying the next John Walker Lindh, Timothy McVeigh or Richard Reid.

The SPOT program tries to identify a specific menu of behaviors that will naturally emerge due to elevated levels of anxiety or stress. The hypothesis is that terrorists would display those cues when attempting to enter a secure facility such as an airport. But behavioral scientists do not agree on these non-verbal cues and they do not agree on whether terrorists would exhibit them. Because it is impossible to get a group of terrorists to participate in a double-blind experiment, it is hard to validate the theory. DHS points to the program's success in identifying people who have violated the law, and are caught, but no one can be certain criminals and terrorists behave in a similar fashion.

TSA relies on non-verbal cues to help sort through the more than I million passengers that fly in the U.S. each day. Non-verbal cues provide a filtering method to allow officers to determine who they should engage in discussion looking for verbal signs of deception. There does is more agreement among social scientists that verbal interactions with individuals can help in detecting deception.

We had hoped that a DRS-funded "validation report" on the SPOT program would be available for this hearing today. That report purportedly shows that SPOT-trained behavior detection officers are much more likely to identify what TSA deems "high risk" passengers as against a purely random sample of passengers. We look forward to the report's completion and its findings; without it we are missing an important initial assessment of the program's performance.

Over the past ten years, since the 9.11 terrorist attacks, Congress has allocated billions of dollars to the Department of Homeland Security for the development of tools and technologies to keep our air travel secure. Too often that investment has been wasted. Too often we have relied on technology that is not adequately tested before it is deployed, is not based upon adequate scientific evidence of its effectiveness and almost inevitably the technology has proven costly to acquire, deploy and service. This Subcommittee has examined some of these DRS technologies in the past, including the Advanced Spectroscopic Portal (ASP) radiation monitors. DRS has been forced to withdraw other technologies and to re-scope and re-think programs, including the ASP program, SBInet, explosive detection "air puffers" and Advanced Imaging Technology (AIT) to screen passengers.

Costing more than $200 million per year we need to carefully evaluate SPOT's operational merit. Is the SPOT program –as it is now constructed worthwhile? Should it be restructured? Should it be expanded? Can it be improved–and if so, how? What are the ultimate costs of the program and would that money be spent elsewhere for greater effect helping to improve security on unsecured non-aviation transportation modes, for instance?

I hope our witnesses can help address some of these issues today. I again want to express my disappointment at the lack of cooperation of TSA with the Committee. One of the reasons that it is unclear to me what training TSA provides BDOs regarding "racial profiling" in their SPOT program is because TSA has so far refused to permit Subcommittee staff to observe this training. They have also refused to provide a witness for this hearing. It is hard to make the case that the SPOT program is working and worthy of continued Congressional funding and support when the agency that runs the program refuses to participate in a hearing. I hope that the agency will rethink their position. I want to thank the Chairman for calling this hearing and I look forward to hearing the testimony of the witnesses who are here today.

Chairman BROUN. Thank you, Ms. Edwards. If there are Members who wish to submit additional opening statements, those statements will be added to the record at this point.

At this time I would like to introduce our panel of our witnesses. Mr. Stephen Lord is the GAO executive responsible for directing GAO's numerous engagements on aviation and service transportation issues. Before his appointment to the Senior Executive Service in 2007, Mr. Lord led GAO's work on a number of key international security, finance, and trade issues. Mr. Lord has received numerous GAO awards for meritorious service, outstanding achievement, and teamwork. Congratulations.

Mr. Larry Willis is the Program Director for suspicious behavior detection within the Human Factors Division of the Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security. Boy, your business card must be a big one with all that.

Detective Lieutenant Peter J.—how do you pronounce your name, sir?

Mr. DIDOMENICA. DiDomenica.

Chairman BROUN. DiDomenica. Okay. Mine is pronounced Broun. My family either can't spell or can't pronounce, so I am very cognizant of people's pronunciation. Detective Lieutenant Peter J. DiDomenica is employed by the Boston University Policy where he commands the Police Detective Division. Prior to this he served as a Massachusetts State Police Officer, as well as the Director of Security Policy at Boston Logan International Airport, where he developed innovative antiterrorism programs.

Dr. Paul Ekman is Professor Emeritus of Psychology at UCSF and is currently the President of the Paul Ekman Group. He has authored or edited 15 books—wow, you have been busy, sir—and has consulted with federal and local law enforcement and national security organizations. The American Psychological Association identified Dr. Ekman as one of the 100 most influential psychologists of the 20th century. Quite an honor, sir. "Time" Magazine selected him as one of the 100 most influential people of 2009. He is also the Scientific Advisor to the dramatic television series on Fox TV, "Lie to Me," which was inspired by his research. I hope you are getting rich with all that. I love the market system. This is great.

Dr. Maria Hartwig is an Associate Professor in the Department of Psychology at John Jay College of Criminal Justice. She has published research on deception in a number of scientific journals, is on the Editorial Board of Law and Human Behavior. In 2008, Dr. Hartwig received an Early Career Award by the European Association of Psychology and Law for her contributions to psychological research. Congratulations.

Dr. Philip Rubin is the Chief Executive Officer and a Senior Scientist at Haskins Laboratories, a private, nonprofit research institute affiliated with Yale University and the University of Connecticut. In 2010, Dr. Rubin received APA's Meritorious Research Service Commendation. Dr. Rubin is the Chair of the National Academies Board on Behavioral, Cognitive, and Sensory Sciences, and was previously the Chair of the National Research Council Committee on Field Evaluation of Behavioral and Cognitive Sciences Based Methods and Tools for Intelligence and Counterintelligence and a member of the NRC Committee on Developing Metrics for Department of Homeland Security's Science and Technology Research.

Noticeably absent from the witness table is the Transportation Security Administration. TSA was invited to the initial hearing on March 13 that was postponed. They were invited to this hearing several weeks ago. In response to these invitations, DHS has refused to send a TSA representative. On another Committee hearing just yesterday the Department of Homeland Security refused to have a witness sit on a panel with other witnesses. DHS has staked out a claim that I think is intolerable. It is unconscionable that TSA will not send their representative here today to this important hearing on this program that is slated to spend $1.2 billion of the taxpayers' money to talk to us about it, and I find that totally reprehensible.

In a letter to this Committee, DHS sought to detail the Subcommittee's interest, presumably quoting from Rule 10 of the House of Representatives that delineates jurisdiction. In this letter they state "Given the Subcommittee's interest in scientific research, development, and demonstration in projects," Larry Willis, Project Manager for the Hostile Intent Detection Validation Project at DHS's Science and Technology Directorate, "S&T will represent DHS at the aforementioned hearing."

I find it highly presumptuous that DHS thinks it knows our jurisdiction better than we do. It shows their arrogance. I find it appalling. Considering this Committee was formed in 1958 and

played an active role in creating the Department of Homeland Security. While DHS surprisingly cites our black-letter jurisdiction under Rule 10 correctly, they must have stopped reading there. Under Rule 11, the Committee on Science, Space, and Technology is tasked with the responsibility to "review and study on a continuing basis laws, programs, and government activities relating to non-military research and development."

Unless TSA and DHS are arguing that science and research played no role in the development of SPOT program, I see a compelling reason for their attendance here today. The nexus between science and operations is vitally important to understanding how programs were developed, why there are problems, and how they can improve.

If TSA and DHS are, in fact, making a claim that science and research played no role in the formation of the program whatsoever, then this program should be shut down immediately for lacking any scientific basis and being little more than snake oil. If DHS does not value this Committee's role in overseeing the Agency and if TSA does not value S&T's scientific advice, there are a number of legislative options that this Committee could employ to change that impression.

I will also note that DHS has sent Agency officials to testify before this Committee from Customs and Border Protection and the Coast Guard. I find it odd that in this instance TSA would not want to talk about this program. It makes me wonder what they are trying to hide. When DHS is asking for a 9.5 percent increase in the fiscal year 2011 budget request for SPOT, you would think that they could justify that increase to us here in Congress.

Let me be clear. The Administration does not tell Congress how to run its hearings. We will likely return to this issue once again after the validation report is delivered. At that point we may seek TSA's input once again. If that is decided, this Committee may seek more aggressive measures to compel TSA's attendance, including the issuance of a subpoena.

This Committee has not needed to issue a subpoena in almost two decades and has been successful in reaching accommodations with Republican and Democratic administrations. I am hopeful that TSA will determine that they have a valuable contribution to make to this topic in the future so that we do not find it necessary to go down that road.

Now, as our witnesses should note, spoken testimony is limited to five minutes each, if you all would please try to hold it to the five minutes. If you go over a few seconds, then that will be okay. But if you just go on and on, then I may have to tap the gavel so you know please wrap up very quickly. Your written testimony will be included in the record of the hearing. It is the practice of the Subcommittee on Investigations and Oversight to receive testimony under oath. Do any of you have any objections to taking an oath? Any of you? Okay. Let the record reflect that all witnesses were willing to take an oath. They all showed that by nodding their head from side to side indicating no. You also may be represented by counsel. Do any of you have counsel here with you today? No? Okay. Let the record reflect that none of the witnesses have counsel. Now, if you would, please, stand and raise your right hand.

Do you solemnly swear or affirm to tell the whole truth and nothing but the truth, so help you, God?

Let the record reflect that all witnesses participating have taken the oath. Thank you. You all may sit down.

I now recognize our first witness, Mr. Stephen Lord, Director of Homeland Security Justice Issues, Government Accountability Office. Mr. Lord, five minutes.

TESTIMONY OF STEPHEN LORD, DIRECTOR, HOMELAND SECURITY AND JUSTICE ISSUES, GOVERNMENT ACCOUNTABILITY OFFICE

Mr. LORD. Thank you. Chairman Broun, Ranking Member Edwards, and other Members of the Committee, thank you for inviting me here today to discuss TSA's behavior-detection program, also known as SPOT.

Today, I would like to discuss two issues. First, DHS's ongoing efforts to validate the program and second, TSA's efforts to make better use of the information collected through this program. This is an important issue as the Department is currently seeking $254 million in fiscal year 2012 funds, including 350 additional Behavioral Officer positions. And as we reported in May 2010, TSA deployed SPOT to 161 airports across the Nation before completing ongoing validation efforts. Thus, it is still unclear whether behavior and appearance indicators can be used to reliably identify individuals who may pose a threat to the U.S. aviation system. According to TSA, the program was deployed before these efforts were completed to help address potential security threats.

To help ensure the program is based on sound science, our report recommended that TSA and DHS convene an independent panel of experts to review the methodology and results of the ongoing validation effort you mentioned in your opening comments. The good news is DHS agreed with this recommendation. However, as other panel members will note in their statements today, a scientific consensus does not yet exist on whether behavior detection principles can be reliably used for counterterrorism purposes in an airport environment.

It is also important to note that the current DHS validation effort will not answer several important questions. For example, how long can Behavior Detection Officers observe passengers without becoming fatigued? What is the optimal number of officers needed to ensure adequate coverage? To what extent are the behavior and appearance indicators the right mix of indicators? Should the list of indicators be larger or should the list be smaller? Also, while Mr. Willis will report that SPOT is nine times more effective than random screening in identifying so-called high-risk individuals, the results of this analysis have yet to be shared with GAO or independently reviewed.

Our report also highlighted some difficulties that TSA faced in capturing and analyzing the rich information that was collecting at airports. Thus, we recommended that TSA better collect and analyze SPOT information to help connect the dots on passengers who may pose a threat to the U.S. aviation system.

For example, we recommended that TSA clarify its guidance to BDOs for inputting information into the database used to track

suspicious activities. We also recommended that they expand access to this database across all SPOT airports. The good news is TSA agreed with our recommendations and has revised its procedures accordingly. TSA also expanded access to this database to all SPOT airports as of March of this year.

Our 2010 report also recommended that TSA make better use of information collected through airport video systems. We noted that 16 individuals who were later charged with or pleaded guilty to terrorism-related offenses transited through eight SPOT airports on 23 different occasions. Thus, we recommended that TSA examine the feasibility of using airport video systems to refine the current number of behaviors currently assessed and also to use this information to help refine the program going forward. We believe such recordings could help identify behaviors that may be common among terrorists or could demonstrate that terrorists do not generally display any identifying behaviors. Again, TSA agreed with our recommendation and is now exploring ways to better use these video recordings.

In closing, behavior and appearances monitoring might be able to play a useful role in airport counterterrorism efforts. However, it is still an open question whether these techniques can be successfully applied on a large scale in the airport environment. And while I am encouraged that DHS has taken steps to validate the program, I am still surprised the Department is seeking additional funding for this program before the issue is fully addressed. Now, hopefully, today's hearing will help clarify S&T's future plans for validating the program.

Chairman Broun, Ranking Member Edwards, and other Members of the Committee, this concludes my statement. I look forward to your questions.

[The prepared statement of Mr. Lord follows:]

PREPARED STATEMENT OF MR. STEPHEN LORD, DIRECTOR, HOMELAND SECURITY AND JUSTICE ISSUES, GOVERNMENT ACCOUNTABILITY OFFICE

United States Government Accountability Office

GAO

Testimony

Before the Subcommittee on Investigations and Oversight, Committee on Science, Space, and Technology, House of Representatives

For Release on Delivery
Expected at 10:00 a.m. EDT
Wednesday, April 6, 2011

AVIATION SECURITY

TSA Is Taking Steps to Validate the Science Underlying Its Passenger Behavior Detection Program, but Efforts May Not Be Comprehensive

Statement of Stephen M. Lord, Director
Homeland Security and Justice Issues

GAO
Accountability * Integrity * Reliability

GAO-11-461T

April 6, 2011

GAO
Accountability • Integrity • Reliability

Highlights

Highlights of GAO-11-461T, a testimony before the Subcommittee on Investigations and Oversight, Committee on Science, Space, and Technology, House of Representatives

AVIATION SECURITY

TSA Is Taking Steps to Validate the Science Underlying Its Passenger Behavior Detection Program, but Efforts May Not Be Comprehensive

Why GAO Did This Study

The attempted passenger aircraft bombing of Northwest flight 253 on December 25, 2009, provided a vivid reminder that the civil aviation system remains an attractive terrorist target. To enhance aviation security, in October 2003 the Department of Homeland Security's (DHS) Transportation Security Administration (TSA) began testing of its Screening of Passengers by Observation Techniques (SPOT) program to identify persons who may pose a risk to aviation security. The SPOT program utilizes behavior observation and analysis techniques to identify potentially high-risk passengers. This testimony provides information on (1) the extent to which TSA has validated the scientific basis for SPOT and (2) other operational challenges. This statement is based on a prior report GAO issued in May 2010 on SPOT, including selected updates made in March 2011. For the updates, GAO reviewed documentation on TSA's progress in implementing the report's recommendations.

What GAO Recommends

GAO has made recommendations in prior work to strengthen TSA's SPOT program. TSA generally concurred with the recommendations and has actions under way to address them. GAO provided the updated information to TSA. TSA had no comment.

View GAO-11-461T or key components.
For more information, contact Stephen M. Lord at (202) 512-4379 or lords@gao.gov

What GAO Found

As GAO reported in May 2010, TSA deployed its behavior detection program nationwide before first determining whether there was a scientifically valid basis for the program. According to TSA, the program was deployed before a scientific validation of the program was completed in response to the need to address potential security threats. However, a scientific consensus does not exist on whether behavior detection principles can be reliably used for counterterrorism purposes, according to a 2008 report of the National Research Council of the National Academy of Sciences. DHS is conducting a study on the scientific basis of SPOT. Thus, in May 2010, GAO recommended that DHS convene an independent panel of experts to review the methodology of its study. DHS concurred and stated that it is convening an independent panel to review its current efforts to help validate the scientific basis for the program, which is expected to complete its work by early April 2011. Nonetheless, DHS's study to assess SPOT is not designed to fully validate whether behavior detection can be used to reliably identify individuals in an airport environment who pose a security risk. For example, factors such as the length of time behavior detection officers (BDO) can observe passengers without becoming fatigued are not part of the plan and could provide additional information on the extent to which SPOT can be effectively implemented. The results of a panel to review DHS's methodology could help ensure a rigorous, scientific validation of SPOT.

As GAO previously reported, TSA experienced SPOT operational challenges, including not systematically collecting and analyzing information obtained by BDOs on passengers who may pose a threat to the aviation system. Better utilizing existing resources would enhance TSA's ability to quickly verify passenger identity and could help TSA to more reliably "connect the dots" with regard to persons who pose a threat. Thus, GAO recommended that TSA clarify BDO guidance for inputting information into the database used to track suspicious activities, and develop a schedule to expand access to this database across all SPOT airports. TSA agreed and in March 2011 stated that it has revised the SPOT standard operating procedures on how BDOs are to input data into the database used to report suspicious activities. TSA plans to implement these revised procedures in April 2011. TSA also reported that all SPOT airports have access to this database as of March 2011. In addition, GAO reported that individuals allegedly involved in six terrorist plots transited SPOT airports. GAO recommended in May 2010 that TSA study the feasibility of using airport video recordings of the behaviors exhibited by persons transiting airport checkpoints who were later charged with or pleaded guilty to terrorism-related offenses. GAO reported that such recordings could provide insights about behaviors that may be common among terrorists or could demonstrate that terrorists do not generally display any identifying behaviors. TSA agreed that studying airport videos could be a useful tool in understanding terrorist behaviors in the airport environment and in March 2011 reported that it is exploring ways to better utilize such recordings.

Chairman Broun, Ranking Member Edwards, and Members of the Subcommittee:

I appreciate the opportunity to participate in today's hearing to discuss the Transportation Security Administration's (TSA) behavior-based passenger screening program known as the Screening of Passengers by Observation Techniques (SPOT) program. The attempted U.S. passenger aircraft bombing of Northwest flight 253 on December 25, 2009, provided a vivid reminder that civil aviation remains an attractive terrorist target and underscores the need for effective passenger screening. To help enhance aviation security, in October 2003, the Department of Homeland Security's (DHS) TSA began testing its SPOT program to identify persons who may pose a risk to aviation security. The SPOT program utilizes behavior observation and analysis techniques to identify potentially high-risk passengers. TSA designed SPOT to provide behavior detection officers (BDO) with a means of identifying persons who may pose a potential security risk at TSA-regulated airports by focusing on behaviors and appearances that deviate from an established baseline and that may be indicative of stress, fear, or deception.

In instances when a passenger's SPOT indicators place him or her above a numerical threshold, he or she will be directed to the second step of SPOT, referral screening. This involves additional questioning and physical search of his or her person and property by BDOs and transportation security officers. This referral screening occurs in the checkpoint area. A referral to a law enforcement officer (LEO) is a potential third step in the SPOT process. After a passenger has been referred by the BDOs to a LEO, the LEO is then expected to independently determine, through additional investigation, such as questioning the passenger and, if appropriate, conducting an identity verification and background check through the Federal Bureau of Investigation's (FBI) National Crime Information Center (NCIC), whether sufficient grounds exist to take further action, such as detaining or arresting the passenger. BDOs have been selectively deployed to 161 of the 462 TSA-regulated airports in the United States. The conference report accompanying the fiscal year 2010 DHS appropriations act provided $211.9 million for the SPOT program.[1] The administration has requested $232 million for SPOT for fiscal year 2011, a $20.2 million (9.5 percent) increase over the fiscal year 2010 funding level, to support 3,350 BDOs. If this increase is appropriated, TSA will have invested over $800

[1] H.R. Rep. 111-298, at 77 (2009) (Conf. Rep.).

million in the program since fiscal year 2007. In addition, DHS has requested about $254 million, a $21.9 million increase, in fiscal year 2012 to support an additional 350 BDOs.

My statement today discusses TSA's and DHS's efforts to validate the scientific basis of the SPOT program, as well as steps that TSA is taking to address operational challenges in deploying SPOT to airports. My comments are based primarily on our May 2010 report.[2] It also includes selective updates we obtained in March 2011. For our May 2010 report, we reviewed relevant literature on behavior analysis by subject matter experts. This included a 2008 study by the National Research Council of the National Academy of Sciences that included a discussion section on the issue of deception and behavioral surveillance, as well as other issues related to behavioral analysis.[3] We interviewed recognized experts in the field, as well as cognizant officials from other U.S. government agencies that utilize behavior analysis in their work, including U.S. Customs and Border Protection (CBP), the U.S. Secret Service, the Federal Air Marshal Service (FAMS), and the FBI. To better understand how SPOT incorporated expertise on behavior analysis for aviation security, we also interviewed current and retired officials of Israel's El Al Airlines,[4] whose

[2] See GAO, *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 (Washington, D.C.: May 20, 2010).

[3] National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008). The report's preparation was overseen by the National Academy of Sciences Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals. Although the report addresses broader issues related to privacy and data mining, a senior National Research Council official stated that the committee included behavior detection as a focus because any behavior detection program could have privacy implications.

[4] Although SPOT is based in some respects on El Al's aviation security program, El Al's processes differ in substantive ways from those used by the SPOT program. In particular, El Al does not use a list of specific behaviors with numerical values for each, or a numerical threshold to determine whether to question a passenger; rather, El Al security officers utilize behavioral indicators as a basis for interviewing all passengers boarding El Al passenger aircraft, and access relevant intelligence databases, when deemed appropriate. According to these officials, El Al also permits what is termed "profiling," in which passengers may be singled out for further questioning based on their nationality, ethnicity, religion, appearance, or other descriptive characteristics, but these are not the only bases on which a passenger may be questioned. The scale of El Al operations is considerably smaller than that of major airlines operating within the United States. In Israel, El Al operates out of one hub airport; in contrast, there are 462 TSA-regulated airports in the United States.

security processes TSA cites as providing part of the basis of the SPOT program. To identify any challenges that emerged during implementation of the SPOT program, we conducted field site visits to 15 TSA-regulated airports with SPOT, which represent almost 10 percent of the 161 TSA-regulated airports with SPOT to observe operations and meet with key program personnel. To obtain comparative data on how SPOT had been implemented at different airports across the nation, we conducted a survey of all federal security directors responsible for security operations at TSA-regulated airports with SPOT. We obtained a 100 percent response rate. In addition, to determine if individuals who were later charged with or pleaded guilty to terrorism-related offenses had transited SPOT airports and whether TSA could obtain information from these transits to enhance its understanding of terrorist behaviors, we reviewed CBP and Department of Justice information to (1) identify individuals who were charged with or pleaded guilty to terrorism-related offenses and (2) determine if these individuals had, prior to being charged, transited airports where SPOT had been deployed. For the updates, we reviewed documentation from TSA on the steps it has taken to implement the recommendations from our May 2010 report. More detailed information about our scope and methodology is included in our May 2010 report. We conducted this work in accordance with generally accepted government auditing standards.

TSA Did Not Validate the Science Underlying the SPOT Program before Deploying SPOT

As discussed in our May 2010 report, TSA deployed SPOT nationwide before first determining 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. A validation study by DHS's Science and Technology Directorate is under way now, but questions exist regarding whether the study's methodology is sufficiently comprehensive to validate the SPOT program. Specifically, DHS's plan to assess SPOT is not designed to fully validate whether behavior detection can be used to reliably identify individuals in an airport environment who pose a security risk. The results of an independent assessment are needed to determine whether current validation efforts are sufficiently comprehensive to validate the program, and to support future requests for increased funding.

According to TSA, SPOT was deployed before a scientific validation of the program was completed, but TSA stated that this deployment was made in response to the need to address potential threats to the aviation system, such as suicide bombers. TSA also stated that the program was based upon scientific research available at the time regarding human behaviors.

Moreover, TSA stated that no other large-scale U.S. or international screening program incorporating behavior- and appearance-based indicators has ever been rigorously scientifically validated.

However, a 2008 report issued by the National Research Council of the National Academy of Sciences stated that the scientific evidence for behavioral monitoring is preliminary in nature.[5] The report also noted that an information-based program, such as a behavior detection program, should first determine if a scientific foundation exists and use scientifically valid criteria to evaluate its effectiveness before deployment. The report added that such 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.[6]

As we reported in May 2010, 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. Thus, we recommended that the Secretary of Homeland Security convene an independent panel of experts to review the methodology of the validation study on the SPOT program being conducted to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program. We also recommended that this assessment include appropriate input from other federal agencies with expertise in behavior detection and relevant subject matter experts.[7] DHS concurred and stated that its current validation study includes an independent review of the study that will include input from a broad range of federal and operational agencies and relevant experts, including those from academia. According to DHS's Science and Technology Directorate, this independent review is expected to be completed in early April 2011.

[5] Specifically, the report states that the scientific support for linkages between behavioral and physiological markers and mental state is strongest for elementary states, such as simple emotions; weak for more complex states, such as deception; and nonexistent for highly complex states, such as when individuals hold terrorist intent and beliefs.

[6] A study performed by the JASON Program Office raised similar concerns. The JASON Program Office is an independent scientific advisory group that provides consulting services to the U.S. government on matters of defense science and technology.

[7] See GAO-10-763.

As discussed in our May 2010 report, DHS has contracted with the American Institutes for Research to conduct its validation study. DHS stated that the ongoing independent review will include, among other things, recommendations on additional studies that should be undertaken to more fully validate the science underlying the SPOT screening process. As we noted in our report, research on other issues, such as determining the number of individuals needed to observe a given number of passengers moving at a given rate per day in an airport environment or the duration that such observation can be conducted by BDOs before observation fatigue affects effectiveness, could provide additional information on the extent to which SPOT can be effectively implemented in airports. Additional research could also help determine the need for periodic refresher training for the BDOs since research has not yet determined whether behavior detection is easily forgotten or can be potentially degraded with time or lack of use. Because such questions exist, the results of an independent panel of experts to assess the methodology of the study could provide DHS with additional assurance regarding whether the study's methodology is sufficiently comprehensive to validate the SPOT program.

Moreover, DHS stated that its current effort to validate the science underlying SPOT includes 3 years of operational SPOT referral data and preliminary results indicate that it is supportive of SPOT. However, in May 2010, we reported weaknesses in TSA's process for maintaining operational data from the SPOT program database. Because of these data-related issues, we reported that meaningful analyses could not be conducted to determine if there is an association between certain behaviors and the likelihood that a person displaying certain behaviors would be referred to a law enforcement officer or whether any behavior or combination of behaviors could be used to distinguish deceptive from nondeceptive individuals.[8]

As we reported in March 2011, Congress may wish to consider limiting program funding pending receipt of an independent assessment of TSA's SPOT program.[9] We identified potential budget savings of about $20 million per year if funding were frozen at current levels until validation efforts are complete. Specifically, in the near term, we reported that

[8] See GAO-10-763.

[9] See GAO, *Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue*, GAO-11-318SP (Washington, D.C.: Mar. 1, 2011).

Congress could consider freezing appropriation levels for the SPOT program at the 2010 level until the validation effort is completed. Assuming that TSA is planning to expand the program at a similar rate each year, this action could result in possible savings of about $20 million per year, or $100 million over 5 years, since TSA is seeking about a $20 million increase for SPOT in fiscal year 2011. We also reported that upon completion of the validation effort, Congress may also wish to consider the study's results—including those on the program's effectiveness in using behavior-based screening techniques to detect terrorists in the aviation environment—in making future funding decisions regarding the program.

TSA Is Taking Steps to Address Operational Challenges in Implementing the SPOT Program

In May 2010, we reported that TSA is not fully utilizing the resources it has available to systematically collect the information obtained by BDOs on passengers whose behaviors and appearances resulted in either a referral to a BDO or to a LEO, and who thus may pose a risk to the aviation system. As we previously reported, TSA does not provide official guidance on how or when BDOs or other TSA personnel should enter data into the Transportation Information Sharing System or which data should be entered.[10] Official guidance on what data should be entered into the system on passengers could better position TSA personnel to be able to consistently collect information to facilitate synthesis and analysis in "connecting the dots" with regard to persons who may pose a threat to the aviation system.

Moreover, as of May 2010, TSA had not developed a schedule or milestones by which database access would be deployed to SPOT airports, or a date by which access at all SPOT airports would be completed. Setting milestones for expanding Transportation Information Sharing System access to all SPOT airports, and setting a date by which the expansion will be completed, could better position TSA to identify threats to the aviation system that may otherwise go undetected and help TSA track its progress in expanding Transportation Information Sharing System access as management intended. Thus, we previously

[10] The Transportation Information Sharing System is a database owned by TSA's FAMS component; the data entered into it may be shared with other federal, state, or local law enforcement and law enforcement support entities. Federal air marshals file reports related to the observation of suspicious activities and input this information, as well as incident reports submitted by airline employees and other individuals within the aviation domain, into the Transportation Information Sharing System.

recommended that TSA provide guidance in the SPOT standard operating procedures or other directives to BDOs, and to other TSA personnel as appropriate, on how and when to input data into the Transportation Information Sharing System database.[11] In March 2011, TSA stated that it has taken steps to implement our recommendation by revising SPOT standard operating procedures to provide guidance directing the input of BDO data into the Transportation Information Sharing System. TSA plans to implement these revised procedures in April 2011. In addition, all SPOT airports have access to the Transportation Information Sharing System as of March 2011 according to TSA.

In addition, as we previously reported, studying airport video recordings of the behaviors exhibited by persons transiting airport checkpoints who were later charged with or pleaded guilty to terrorism-related offenses could provide important insights about behaviors that may be common among terrorists or could demonstrate that terrorists do not generally display any identifying behaviors. In addition, such images could help determine if BDOs are looking for the right behaviors or seeing the behaviors they have been trained to observe.

Using CBP and Department of Justice information, we examined the travel of key individuals allegedly involved in six terrorist plots that have been uncovered by law enforcement agencies.[12] We determined that at least 16 of the individuals allegedly involved in these plots moved through 8 different airports where the SPOT program had been implemented.[13] Six of the 8 airports were among the 10 highest-risk airports, as rated by TSA in its Current Airport Threat Assessment. In total, these individuals moved through SPOT airports on at least 23 different occasions. For example, according to Department of Justice documents, in December 2007 an

[11] See GAO-10-763.

[12] The analysis included only flights leaving the United States. Department of Justice data show that more than 400 individuals have been convicted in the United States for terrorism-related offenses since September 11, 2001. We did not examine the travel itineraries of all these individuals.

[13] The events included the Mumbai, India, attack of 2008; a plot to attack the Quantico, Virginia, Marine base in 2008; an effort by five Americans to receive training and fight in Pakistan in December 2009; a plot to attack infrastructure in New York City in 2009; an effort to provide men and support for terrorists in Somalia in 2008; and an attack on a U.S. base in Afghanistan by an American who received training in Pakistan. We were unable to confirm whether BDOs were stationed at the checkpoints used by these individuals at the time they traveled.

individual who later pleaded guilty to providing material support to Somali terrorists boarded a plane at the Minneapolis-Saint Paul International Airport en route to Somalia. Similarly, in August 2008, an individual who later pleaded guilty to providing material support to al Qaeda boarded a plane at Newark Liberty International Airport en route to Pakistan to receive terrorist training to support his efforts to attack the New York subway system.

Our survey of federal security directors at 161 SPOT airports indicated that most checkpoints at SPOT airports have surveillance cameras installed. Thus, we reported that TSA may be able to utilize the information collected from the video infrastructure at the nation's airports to study the behavior of persons who were later charged with or pleaded guilty to terrorism-related offenses to help improve and refine the existing SPOT program. As a result, in our May 2010 report, we recommended that if the current validation effort determines that the SPOT program has a scientifically validated basis for using behavior detection for counterterrorism purposes in the airport environment, then TSA should study the feasibility of using airport checkpoint surveillance video recordings to enhance its understanding of terrorist behaviors.[14] DHS agreed with our recommendation and noted that TSA agrees this could be a useful tool and is working with DHS's Science and Technology Directorate to utilize video case studies of terrorists, if possible. TSA officials agreed that examining video recordings of individuals who were later charged with or pleaded guilty to terrorism-related offenses, as they used the aviation system to travel to overseas locations allegedly to receive terrorist training or to execute attacks, could help inform the SPOT program's identification of behavioral indicators. In March 2011, TSA stated that it is exploring ways to better utilize video recordings to identify these behavioral indicators.

Chairman Broun, Ranking Member Edwards, and Members of the Subcommittee, this concludes my statement. I look forward to answering any questions that you may have at this time.

[14] See GAO-10-763.

Contacts and Acknowledgments

For questions about this statement, please contact Stephen M. Lord at (202) 512-4379 or lords@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this statement. Individuals making key contributions to this testimony are David M. Bruno, Assistant Director; Ryan Consaul; Katherine Davis; Emily Gunn; and Tracey King.

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.

GAO's Mission	The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.
Obtaining Copies of GAO Reports and Testimony	The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday afternoon, GAO posts on its Web site newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to www.gao.gov and select "E-mail Updates."
Order by Phone	The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's Web site, http://www.gao.gov/ordering.htm. Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537. Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.
To Report Fraud, Waste, and Abuse in Federal Programs	Contact: Web site: www.gao.gov/fraudnet/fraudnet.htm E-mail: fraudnet@gao.gov Automated answering system: (800) 424-5454 or (202) 512-7470
Congressional Relations	Ralph Dawn, Managing Director, dawnr@gao.gov, (202) 512-4400 U.S. Government Accountability Office, 441 G Street NW, Room 7125 Washington, DC 20548
Public Affairs	Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800 U.S. Government Accountability Office, 441 G Street NW, Room 7149 Washington, DC 20548

Please Print on Recycled Paper

Chairman BROUN. Thank you, Mr. Lord. I now recognize our next witness, Dr. Paul Ekman, Professor Emeritus—wait a minute. I skipped over one and I apologize. I now recognize Mr. Willis—our next witness, Mr. Larry Willis, Program Manager, Homeland Security Advanced Research Project Agency, Science and Technology Directorate, Department of Homeland Security. Mr. Willis, you have five minutes. Thank you, sir.

TESTIMONY OF LARRY WILLIS, PROGRAM MANAGER, HOMELAND SECURITY ADVANCED RESEARCH PROJECTS AGENCY, SCIENCE AND TECHNOLOGY DIRECTORATE, DEPARTMENT OF HOMELAND SECURITY

Mr. WILLIS. Thank you. Good afternoon, Chairman Broun, Ranking Member Edwards, distinguished Members of the Subcommittee. I am honored to appear before you today on behalf of the Department of Homeland Security, Science and Technology Directorate, to discuss our evaluation of the Transportation Security Administration's Screening Passenger by Observation Technique, or SPOT referral report, which is a checklist of predefined behavior indicators used by TSA to identify potentially high-risk travelers.

For the purpose of S&T's studies, high-risk travelers are defined as those passengers in possession of serious prohibited and/or illegal items or individuals engaging in conduct leading to arrest.

For background purposes, the SPOT validation effort began in 2007 as a result of the component-led, S&T-managed People Screening Capstone Integrated Product Team process that identified and prioritized capability gaps of DHS operational customers. As an active participant in this IPT process, TSA identified the SPOT Referral Report and its associated indicators as a candidate for the validation study. The SPOT Referral Report contains a discrete list of observable indicators which have been designated by TSA as Sensitive Security Information, or SSI. TSA's Behavior Detection Officers, or BDOs, are trained to identify these indicators and use them to make screening decisions, such as referral for additional screening at the TSA checkpoint.

It is important to note that the behavioral screening isn't limited to aviation security and is conducted formally or informally by DHS agencies, the Department of Defense, the intelligence community, and law enforcement worldwide. The SPOT validation research is a rigorous evaluation of TSA's SPOT Referral Report that supports our better understanding of the threat, the screening accuracy of the existing indicators, and advances of science of behavioral-based screening.

S&T, in cooperation with the American Institute for Research designed the Base Rate Study to compare TSA's SPOT Referral Report process with a random screening process. AIR is one of the largest non-profit behavioral science research organizations in North America and has performed numerous validation studies. Two databases were used for the study.

The first was designed to include case information from randomly selected travelers who were subjected to the SPOT referral process during the Base Rate Study conducted from December 2009 through October 2010 and included a total of 71,589 referrals from 43 airports. To make direct comparisons between the Base Rate

database and the Operational Referrals, a second dataset was created for the 23,265 Operational SPOT Referrals collected during the same time and at the same locations of the Base Rate Study.

Together, these two datasets allowed AIR to assess the extent to which the SPOT Referral Report of observable indicators lead to correct screening decisions. A key number of findings emerged from the analysis of the SPOT Referral Report, including the following, which I would like to share with you.

One, Operational SPOT identifies high-risk travelers at a significantly higher rate than random screening. The study data indicate that a high-risk traveler is nine times more likely to be identified using Operational SPOT versus random screening. Moreover, to achieve this outcome, BDOs within the study were able to engage 50,000 fewer travelers using Operational SPOT than with random selection methods.

The second result is a population base rate for SPOT indicators is low. Among those selected for random screening the Base Rate Study, the most frequently observed indicator was displayed in only 2.8 percent of the randomly selected travelers. All of the other indicators were observed in fewer than two percent of the travelers selected during the Base Rate Study.

In conclusion, these results indicate that the SPOT program is significantly more accurate than random screening in identifying high-risk travelers using the metrics that we employed. Our validation process, which included an independent and comprehensive review of SPOT Referral Report, is a key example of how S&T works to enhance the effectiveness of the Department's operational activities.

Chairman Broun, Ranking Member Edwards, I thank you again for this opportunity to discuss the research to validate the Screening of Passengers by Observation Technique Referral Report. And I am happy to answer the questions that the Subcommittee may have.

[The prepared statement of Mr. Willis follows:]

PREPARED STATEMENT OF MR. LARRY WILLIS, PROGRAM MANAGER FOR THE SCIENCE AND TECHNOLOGY DIRECTORATE, DEPARTMENT OF HOMELAND SECURITY

Introduction and Study Objective:

Good afternoon, Chairman Broun, Ranking Member Edwards and distinguished Members of the Subcommittee. I am honored to appear before you today on behalf of the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) to discuss our evaluation of the Transportation Security Administration's (TSA) Screening of Passengers by Observation Techniques (SPOT) program. SPOT is a behavior observation and analysis program in which personnel are trained to identify behaviors that deviate from an established baseline that could be possible indicators for terrorism or criminal activity. Today, I will describe S&T's research assessing the validity of the SPOT Referral Report, which is a checklist of predefined observable indicators used by TSA to identify potentially high risk travelers. For the purpose of S&T's study, high risk travelers are defined as those passengers in possession of serious prohibited and/or illegal items or individuals engaging in conduct leading to an arrest. Specifically, our study offers an assessment of the extent to which the SPOT Referral Report of observable indicators leads to correct screening decisions at the security checkpoint.

Research Requirements and Background:

Approximately 1.2 million people fly within the United States daily. The SPOT program trains TSA personnel to serve as an additional layer of security in airports by providing a non-intrusive means of identifying individuals who may pose a risk of terrorism or criminal activity. In behavior-based screening, trained personnel at-

tempt to identify anomalous behaviors by observing passengers and comparing what they see to an established behavioral baseline of other passengers developed in the same general location and within the same timeframe. It is important to note that behavioral screening isn't limited to aviation security and is conducted formally or informally by other DHS agencies, the Department of Defense, the Intelligence Community, and law enforcement worldwide. The SPOT validation effort appears to be the most rigorous evaluation of behavioral-based screening.

The SPOT validation effort began in 2007 as a result of the component-led, S&T-managed People Screening Capstone Integrated Product Team (IPT) process that identified and prioritized capability gaps of DHS operational components.

The "People Screening" Capstone IPT established the research requirement to identify and validate observable behavior indicators of threats and suspicious behaviors in a screening environment. As an active participant in this IPT, TSA identified the SPOT Referral Report and its associated indicators as a candidate for the validation study. Through a series of interactions with TSA, S&T determined that the SPOT screening process and the effectiveness of the observable indicators list was testable. The SPOT Referral Report contains a discrete list of observable indicators which have been designated by TSA as Sensitive Security Information (SSI). TSA's Behavior Detection Officers (BDOs) are trained to identify these indicators and use them to make screening decisions, such as referral for additional screening at the TSA checkpoint. Furthermore, TSA records each behavior-based screening event, as well as its corresponding indicators, screening results, and outcomes to help inform future screening decisions. The SPOT process leads to three possible actions: the traveler proceeds through the TSA checkpoint and to their flight as normal; the traveler is identified as possibly carrying serious prohibited/illegal items and receives additional screening at the TSA checkpoint; or the traveler is identified to a Law Enforcement Officer (LEO) for appropriate intervention.

Research Approach:

S&T, in cooperation with the American Institutes for Research (AIR), designed the Base Rate Study to compare TSA's SPOT Referral Report process with a random screening process and to estimate the population base rate of high-risk travelers. AIR is one of the largest non-profit behavioral science research organizations in North America and has performed numerous validation studies. Two databases were used for this study. The first was designed to include case information from randomly selected travelers who were subjected to the SPOT referral process during the Base Rate Study from December 1, 2009 through October 31, 2010, including a total of 71,589 referrals from 43 airports. To make direct comparisons between the Base Rate database and the Operational SPOT Referrals, a second dataset (SPOT comparison dataset) was extracted from TSA's SPOT Referral database to contain the 23,265 Operational SPOT referrals collected during the same time period and from locations covered by the Base Rate Study. Together, these two datasets allowed AIR to assess the extent to which the SPOT Referral Report of observable indicators leads to correct screening decisions at the security checkpoint.

Research Results:

A number of key findings emerged from the analysis of the SPOT Referral Report, including four that I would like to share with you:

1. Operational SPOT identifies high-risk travelers at a significantly higher rate than random screening. The study data indicate that a high risk traveler is nine times more likely to be identified using Operational SPOT versus random screening. (Operational SPOT refers to the standard operating procedure of the BDOs executing the referral reporting process at the checkpoint as opposed to the program as a whole.) Moreover, to achieve these outcomes, BDOs were able to engage with 50,000 fewer travelers using Operational SPOT than they did when using random selection methods.
2. SPOT indicators appear to be observed and utilized consistently across varying airport characteristics. When we examined the consistency in implementation overall, we found that observable indicators within the SPOT Referral Report are used at relatively the same rate regardless of the year, time of year, or size of airport. Moreover, indicators tended to be consistently related to outcomes in the same ways across these characteristics, providing further evidence that the indicators are reliable. These results also serve as initial support for reliability in the use of the SPOT Referral Report, with little to no evidence of major coding variations or random fluctuations.
3. The population base rate for high-risk travelers is extremely low. In other words, the large majority of travelers pose no security risks. Results of the

Base Rate Study confirm that the measurable outcomes that represent high-risk travelers are rare events. These data indicate that the estimated population parameter for:

i. Arrested by Law Enforcement Officer is 1 in 10,000 travelers (or 0.01 percent).

ii. Possession of Fraudulent Documents is 1 in 2,000 travelers (or 0.05 percent).

iii. Possession of Serious Prohibited/Illegal Items is 1 in 750 travelers (or 0.13 percent).

iv. Combined Outcome, or presence of any outcome (of the above), is 1 in 750 travelers (or 0.13 percent).

4. The population base rate for SPOT indicators is low. Among those selected for random screening in the Base Rate Study, very few travelers (approximately 8 percent) exhibited any SPOT indicators. The most frequently observed indicator (again, SPOT indicators are designated SSI) was displayed in only 2.8 percent of the randomly selected travelers. In contrast, this indicator is exhibited in more than half of SPOT-referred travelers. All of the other indicators were observed in fewer than 2 percent of the travelers selected by the Base Rate Study.

Conclusion:

In conclusion, these results indicate that the SPOT program is significantly more effective than random screening: a high-risk traveler is nine times more likely to be identified using Operational SPOT versus random screening. Our validation process, which included an independent and comprehensive review of SPOT, is a key example of how S&T works to enhance the effectiveness of the Department's operational activities. Expanding on these initial findings, we would like to conduct further research to assess the screening accuracy of these observable indicators in similar operational screening environments, in aviation and beyond. Additionally, we would like to work to identify other indicators that could further increase accuracy in operational screening.

Chairman Broun, Ranking Member Edwards, I thank you again for this opportunity to discuss the Screening of Passengers by Observation Techniques program. I am happy to answer any questions the Subcommittee may have.

Chairman BROUN. Thank you, Mr. Willis. You kept your remarks under five minutes, and sometimes that is not done here. In fact, most times it is not done here.

Our next witness is Mr. Peter DiDomenica of the Boston University Police. Thank you, Lieutenant. Appreciate it. You have five minutes, sir.

TESTIMONY OF PETER J. DIDOMENICA, LIEUTENANT DETECTIVE, BOSTON UNIVERSITY POLICE

Mr. DIDOMENICA. Thank you. Good morning. Chairman Broun, Ranking Member Edwards, and Members of the Committee, I thank you for this opportunity to address you today regarding the future of the TSA SPOT program that I originally developed.

By way of additional background, I have trained over 3,000 police, intelligence, and security officials in over 100 federal, state, and local agencies in the United States and U.K. in behavior assessment. I have also been a lecturer or advisor on behavior assessment for the FBI, CIA, Secret Service, DHS, U.S. Army Night Vision Lab, Defense Department Criminal Investigations Task Force, and the National Science Foundation. I appear today representing only myself and not any of the organizations I am or have been employed by.

On December 22, 2001, while assigned to Logan International Airport as a member of the State Police, I was part of a large team of public safety officials who responded to the airfield to meet

American Airlines flight 63, diverted to Boston from a flight from Paris, France to Miami. On board was a passenger named Richard Reid who attempted to detonate an improvised explosive device artfully concealed in his footwear that, if successful, would have killed all 197 passengers and crewmembers aboard. As I stood only a few feet away from Reid, who was now securely in custody in the back of a state police cruiser, it hit me that this man was the real thing, that the threat of another terrorist attack by Al Qaeda would not stop, and that we need to do more, much more, to properly screen passengers than merely focusing on weapons detection. Thus began the development of what would become the Behavior Assessment Screening System or BASS in the SPOT program.

I began to explore the scientific literature in an effort to quantify the human capacity to detect dangerous people. My research included many disciplines including physiology, psychology, neuroscience, as well as specific research into suicide bombers. In developing the program, specific behaviors were selected that were both supported in the scientific literature and consistent with law enforcement experience.

The BASS program went on to be delivered to numerous agencies, including the entire Washington, D.C., Metro Transit Police, Amtrak Police, and the Atlanta Police officers assigned to the world's busiest airport, Atlanta Hartsfield-Jackson International Airport. In 2006, two BASS trainers and I spent two weeks in London where we set up a British version of the BASS program for the British Transport Police as a response to the July 7, 2005, terrorist attacks on the London Underground.

During the course of training police officers around the Nation, the State Police BASS instructors discovered four individuals with suspected terrorist ties. In 2004, while conducting BASS training with the New Jersey Transit Police at Newark Penn Station, I observed three males exhibiting suspicious behavior using BASS techniques. One of the subjects was in the United States on a religious visa from a Middle Eastern country and was being escorted to an Amtrak train for a claimed week-long trip with no luggage. It was later confirmed the subject listed on the visa was on a terror watch list. I even intercepted a DHS inspector on a covert test of the screening checkpoint at Logan Airport in late 2003 with a concealed weapon through BASS techniques.

Although I believe that the SPOT program is effective at identifying high-risk passengers, its effectiveness is limited because proper resolution of highly suspicious people discovered by the TSA BDOs requires a law-enforcement response by police officers trained in the same behavior detection and interview skills. I designed the program so that the most dangerous people would be either removed from the critical infrastructure or arrested by BASS-trained police officers. I do not believe the current TSA airport SPOT familiarization training program is enough. The airport police, in my opinion, need to be trained in the same techniques and skill sets which would engender confidence in the program and their own ability to detect terrorist behavior and prevent additional devastating attacks.

Another issue I see with the SPOT program is that the TSA has created too high an expectation for what it is able to achieve. The

original SPOT program I designed was not primarily for the apprehension of suspects but as a means to deny access to critical infrastructure of high-risk persons who could be involved in terrorism or other dangerous activity. It was to be the last and, most importantly, the best chance to prevent a tragedy when other methods such as intelligence and traditional physical screening have failed. Catching a terrorist through a random encounter in a public place without any prior intelligence is extremely difficult.

By way of example, if we use the known number of terrorist suspects who boarded domestic commercial flights at airports with BDOs and the approximately four billion passenger enplanements at U.S. commercial airports from 2004 to 2009, the base rate of terrorist passengers is about 1 in 173 million. The expectation that the SPOT program will result in the arrest of all terrorists attempting to board a domestic flight in the United States is unrealistic and threatens its continued support. If, however, it is seen as part of a multi-layered approach with the primary goal of preventing terrorist access to critical infrastructure in conjunction with properly trained law enforcement, the program sets reasonable and attainable goals and should have the support of this Congress.

Thank you for this opportunity to address the program and I am prepared to answer any questions that you may have.

[The prepared statement of Mr. DiDomenica follows:]

PREPARED STATEMENT OF MR. PETER J. DIDOMENICA,
LIEUTENANT DETECTIVE, BOSTON UNIVERSITY POLICE

Good morning. Chairman Broun, Ranking Member Edwards, and Members of the Committee, I thank you for this opportunity to address you today regarding the future of the TSA Screening of Passengers by Observation Techniques program that I developed, which is more commonly referred to as the SPOT program.

I am Peter DiDomenica presently employed as a Detective Lieutenant with the Boston University Police Department. I recently joined the Boston University force after serving for more than 22 years with the Massachusetts State Police where I retired as a Lieutenant. While a member of the State Police I served as an investigator in the Major Crime Unit, as the Director of Legal Training for the State Police Academy, as a staff member to five different superintendents, and as Director of Security Policy for Boston Logan International Airport in the two years after the devastating 9/11 attacks. I also served the State Police for a decade as a subject matter expert and lead trainer for Massachusetts police agencies in racial profiling and biased policing. In this capacity I designed statewide police training programs and the State Police traffic stop data collection and analysis system created to monitor enforcement efforts for indications of biased policing. I am also presently a consultant for EOIR Technologies of Fredericksburg, VA where I serve as an advisor on human behavior detection for the U.S. Army Night Vision and Electronic Sensors Directorate. I am a certified instructor in the interview, behavior assessment, and deception detection programs for The Forensic Alliance, a consulting firm of forensic psychologists based in British Columbia, Canada. I am presently an adjunct instructor for the graduate criminal justice program at Anna Maria College in Paxton, MA. I am a licensed attorney in Massachusetts having earned my J.D. in 1995. I have trained over 3,000 police, intelligence, and security officials in over 100 federal, state, and local agencies in the U.S. and U.K. in behavior assessment. I have also been a lecturer or advisor on behavior assessment for the FBI, CIA, Secret Service, Department of Homeland Security, Defense Department Criminal Investigations Task Force, and National Science Foundation. I appear today representing only myself and not any of the organizations I am or have been employed by.

On December 22, 2001, while assigned to Logan International Airport as a member of the State Police and as Director of Security Policy, I was part of a large team of public safety officials who responded to the airfield to meet American Airlines flight 63, diverted to Boston on a flight from Paris, France to Miami. On board was a passenger named Richard Reid who attempted to detonate an improvised explo-

sive device artfully concealed in his footwear that, if successful, would have killed all 197 passengers and crewmembers aboard. As I stood only a few feet away from Reid, who was now securely in custody in the back of a state police cruiser, it hit me that this man was the real thing, that the threat of another terrorist attack from Al Qaeda would not stop, and that we needed to do more, much more, to properly screen passengers than merely focusing on weapons detection. Over the next several days I met with the incident commander for Reid's arrest, Major Tom Robbins, who was the Aviation Security Director for Logan Airport and Troop Commander for State Police Troop F at the airport. One evening, while having dinner with Major Robbins, he wrote the words "walk and talk" on a dinner napkin - a reference to airport narcotics interdiction - and directed me to look into airport drug interdiction programs as a model for a terrorist behavioral profiling program to augment the weapons screening process. Thus began the development of what would become the Behavior Assessment Screening System or BASS.

Because of my legal background and experience in training on racial profiling and bias policing, I knew immediately what the BASS program would not be. Whatever program we would create to identify potential terrorists, it would not include racial profiles that target people of apparent Islamic belief or Arab, Middle Eastern, or South and Central Asian ethnicities. As well as being illegal such profiling could distract security officials from detecting true threats. Moreover, the unconscious bias against these groups would be so strong because of 9/11 that security officials would need training to counter these biases. I began to explore the scientific literature in an effort to quantify the human capacity to detect dangerous people. My research included many disciplines including, physiology, psychology, neuroscience, as well as specific research into suicide bombers. What this literature indicated was that a person who is engaged in a serious deception of consequence or otherwise engaged in an act in which the person has much to lose by being discovered or by failing to succeed will suffer mental stress, fear, or anxiety. Such stress, fear, or anxiety will be manifested through involuntary physical and physiological reactions such as an increase in heart rate, facial displays of emotion, and changes in speed and direction of movement. In developing the program specific behaviors were selected that were both supported in the scientific literature and consistent with law enforcement experience. In addition to avoiding the legal prohibition on selective enforcement based on race, ethnicity, or religion[1] the program also had to ensure that police encounters with the public not meeting the standard of reasonable suspicion were voluntary under the U.S. Supreme Court case of *U.S. v. Medenhall.*[2] In addition to behavior, the program also examines: aspects of appearance unrelated to race, ethnicity, or religion; responses to law enforcement presence and questioning; and, the circumstances surrounding the presence of the person at a specific location. I created a simple method called *"A-B-C-D"* which means *Analysis of Baseline, addition of a Catalyst, and scan for Deviations. Baselines* are merely an evaluation of what was normal for a specific environment and a catalyst is the insertion into the environment of something that would be particularly threatening to a terrorist or criminal to provoke behavioral changes.

In 2002 and 2003 I taught the BASS program to all the troopers, the primary law enforcement agency for Logan Airport, and developed a staff of additional instructors. We also began training other police departments In Massachusetts; in fact we trained the entire Massachusetts Transit Police force and a group of Boston Police officers in preparation for the 2004 Democratic National Convention. Because of the success of the program, I created a derivative program called PASS or the Passenger Assessment Screening System suitable for TSA screeners that eventually became the SPOT program. Over the course of two years I worked with TSA officials at Boston, including the Federal Security Director George Niccara, and officials at TSA headquarters including their Office of Civil Rights, Science and Technology, and Workforce Performance and Training. In 2004 my team of State Police BASS instructors conducted a training program with TSA to create two pilot SPOT programs at Portland International Jetport in Maine and T.F. Green International Airport in Rhode Island.

One of the reasons the BASS program got the interest of TSA headquarters as a model for a behavior detection program was an incident that occurred in the fall of 2003 at Logan Airport while I was training members of the Boston Police in BASS. A middle-age male caught my attention due to an appearance and luggage deviation as well as baseline deviation in movement. When the Boston police officer

[1] Whren v. United States, 517 U.S. 806 at 813 (1996).

[2] 446 U.S. 544 at 554 (1980). ("We conclude that a person has been 'seized' within the meaning of the Fourth Amendment only if, in view of all of the circumstances surrounding the incident, a reasonable person would have believed that he was not free to leave.")

and I engaged this purported passenger in conversation he immediately produced credentials identifying himself as an official of the Department of Homeland Security Office of Investigations and stated he was on his way to test a screening checkpoint to see if they would discover a concealed weapon he was carrying.

The BASS program went on to be delivered to numerous agencies including the entire Washington DC Metro Transit Police, Amtrak Police, and Atlanta Police officers assigned to the world's busiest airport, Atlanta Hartsfield-Jackson International Airport. In 2006 Two BASS trainers and I spent two weeks in London where we set up a British version of BASS for the British Transport Police as a response to the July 7, 2005 terrorist attacks on the London Underground.

During the course of training police officers around the nation, the State Police BASS instructors discovered four individuals with suspected terrorist ties. In 2004, while conducting BASS training with the New Jersey Transit Police at Newark Penn Station, I observed three males exhibiting suspicious behavior using BASS techniques. One of the subjects was in the United States on a religious visa from a Middle Eastern country and was being escorted to an Amtrak train for a claimed week long trip with no luggage. Another subject presented a non-government ID card that was designed to look like a real government ID. There were three behavior cues that led to the encounter followed by three non-verbal cues during the interview as well as conflicting factual statements that made these individuals highly suspicious. It was later confirmed that the subject on the visa was on a terror watch list. In 2004 at the Metro Center rail station in Washington D.C. a member of the BASS training team, while conducting training with the TSA, observed a suspicious male subject who exhibited five behavioral cues under the BASS program. The subject had a British passport with visa stamps from visits to Iraq and was in the U.S. to learn how to fly planes. It was later confirmed that the subject was under investigation for terrorism. Back in 2002 at Logan Airport, a BASS trainer discovered a suspicious subject exhibiting four BASS behavior cues and three non-verbal cues during an interview who had failed to report for deportation and was connected to Ahmed Ressam of the 1999 Millennium bombing plot of Los Angeles Airport.

Unfortunately, since the successful pilot programs in 2004 the TSA has chosen not to continue my services despite my strong recommendation that I remain involved in training, particularly with respect to airport police officers in BASS techniques at airports where the SPOT program is implemented. Although I believe the SPOT program is effective at identifying high risk passengers, its effectiveness is limited because proper resolution of highly suspicious people discovered by the TSA Behavior Detection Officers, or BDOs, requires a law enforcement response by police officers trained in the same behavior detection and interview skills. I designed the program so that the most dangerous people would be either removed from the critical infrastructure or arrested by BASS trained police officers. So, no matter how effective the BDOs are, the most dangerous people will tend to slip through the cracks because of a response by non-BASS trained police officers who may discount the validity of SPOT or who may fail to follow-up with BASS techniques. In most cases where denials of access occur or arrests or detentions are made by police, it is because there are warrants for arrest or because contraband is discovered in the screening process. I do not believe the current TSA airport police SPOT familiarization training program is enough. The airport police, in my opinion, need to be trained in the same techniques and skill sets which will engender confidence in the program and in their own ability to detect terrorist behavior and prevent additional devastating attacks.

Another issue I see with the SPOT program is that the TSA has created too high an expectation for what it is able to achieve. The original SPOT program I designed was not primarily for the apprehension of suspects but as a means to deny access to critical infrastructure of high risk persons who could be involved in terrorism or other dangerous activity. It was to be the last and, most importantly, the best chance to prevent a tragedy when other methods such as intelligence and traditional, needle in the haystack, screening have failed. Catching a terrorist through a random encounter in a public place without any prior intelligence is extremely difficult. By way of example, if we use the number of known terrorism suspects who boarded domestic commercial flights at airports with BDOs, as cited in the Government Accountability Office May 2010 report on Aviation Securitythe last and, most importantly, the best chance to prevent a tragedy when other methods such as intelligence and traditional, needle in the haystack, screening have failed. Catching a terrorist through a random encounter in a public place without any prior intelligence is extremely difficult. By way of example, if we use the number of known terrorism suspects who boarded domestic commercial flights at airports with BDOs, as cited in the Government Accountability Office May 2010 report on Aviation Securitythe last and, most importantly, the best chance to prevent a tragedy when

other methods such as intelligence and traditional, needle in the haystack, screening have failed. Catching a terrorist through a random encounter in a public place without any prior intelligence is extremely difficult. By way of example, if we use the number of known terrorism suspects who boarded domestic commercial flights at airports with BDOs, as cited in the Government Accountability Office May 2010 report on Aviation Security[3], and the approximately 4 billion passenger enplanements at U.S. commercial airports from 2004 to 2009, the base rate of terrorist passengers is about one in every 173 million or .0000006 percent. The expectation that the SPOT program will result in the arrest of all terrorists attempting to board a domestic flight in the United States is unrealistic and threatens its continued support. If, however, it is seen as part of a multi-layered approach with the primary goal of preventing terrorist access to critical infrastructure in conjunction with properly trained law enforcement, the program sets more reasonable and attainable goals.

In 2004 Major Robbins and I, as well as the Massachusetts Port Authority and Massachusetts State Police, were sued by an African-American lawyer for the ACLU who served at the National Coordinator of the American Civil Liberties Union's Campaign Against Racial Profiling. The plaintiff alleged that he was unlawfully detained by the State Police at Logan Airport in October of 2003 and that this unlawful detention was based on BASS training that the troopers received. It was alleged that the BASS training directed the troopers at the airport to detain people without reasonable suspicion of criminal activity and condoned and encouraged racial and ethnic profiling. After a weeklong trial in December 2008 in the Federal District Court for Massachusetts[4], the jury found that the plaintiff was, in fact, unlawfully detained by State Police officers but that the BASS program was not the cause of the unlawful detention. During the trial the judge asked the plaintiff what provisions of the BASS program on its face violate federal law? The plaintiff responded the following provision was unlawful: a provision that allows police, after reasonable efforts to dispel elevated suspicion have failed to escort away from critical infrastructure persons who refuse to identify themselves. The plaintiff also cited the provision allowing for a running of a records check on such persons. The judge ruled from the bench: "I don't see this as on its face being unconstitutional. I mean, there is nothing unconstitutional about running a records check of a person, subjecting a person to additional consensual searches or testing [or] **preventing a person from proceeding into the critical infrastructure or escort[ing] the person away from the critical infrastructure**." (Emphasis added) One of the key components of the BASS program is its anti-detention policy: to empower police to deny persons access to critical infrastructure such as commercial aircraft who display elevated suspicion after reasonable attempts to dispel the suspicion fail. The elevated suspicion is articulable facts and circumstances that do not necessarily have to rise to the level required for a lawful detention under the U.S. Supreme Court case of *Terry v. Ohio*[5]. In keeping with Constitutional mandates, this denial of access in an extremely small number of cases of unresolved suspicion may be the best we can do but it may be enough to prevent a tragedy and it also may provide for the collection of crucial intelligence for an investigation and later arrest. It is important to note that the 9th Circuit U.S. Court of Appeals in the case of Gilmore v. Gonzales has ruled that "the Constitution does not guarantee the right to travel by any particular form of transportation."[6] The Supreme Court has declined to review this decision.

For SPOT to be effective there has to be a cadre of BASS trained police officers to bring about an appropriate resolution from an initial TSA observation. Based on my extensive law enforcement experience using behavioral analysis and those other police officers who have similar experience, as well as having a basic understanding of psychological, neurological, and physiological processes, I know SPOT and BASS techniques do work in identifying potential terrorists and other dangerous people. If done correctly, the process only takes a couple of minutes and is done openly in public areas minimizing interference with the free flow of the public and, most importantly, without interfering with civil rights. This program specifically trains TSA personnel and police officers to counter the effects of unconscious bias that may otherwise result in undue attention on certain ethnic and religious groups and the failure to detect suspicious behavior by truly dangerous people who do not fit the unstated but subconsciously present religious or ethnic profile. When the next shoe bomber or underwear bomber arrives at one of our airports or train stations to blow up one of our planes or subway trains or if they try to gain access to the Super

[3] GAO-10-763. The report cites 23 suspected terrorists having passed through SPOT airports.
[4] King Downing v. Massachusetts Port Authority, et al, Civil Action No. 2004-12513-RBC.
[5] 392 U.S. 1 (1968).
[6] 435 F. 3d 1125.

Bowl or other major sporting event, even when we don't have the constitutional authority to arrest we must have the confidence to deny them access based on the sound principles of BASS and SPOT. This is our last and best chance of preventing another terrorist attack.

Thank you again for this opportunity to address the SPOT program and I am prepared now to answer any questions you may have.

Chairman BROUN. Thank you, Lieutenant. You did not exceed your five minutes either. Congratulations and thank you for being here and——

Mr. DIDOMENICA. Two seconds.

Chairman BROUN. That is right. I recognize our next witness, Dr. Paul Ekman, Professor Emeritus of Psychology, University of California, San Francisco, and President and Founder of the Paul Ekman Group. Doctor, you have five minutes for your testimony.

TESTIMONY OF PAUL EKMAN,
PROFESSOR EMERITUS OF PSYCHOLOGY,
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO,
AND PRESIDENT AND FOUNDER, PAUL EKMAN GROUP, LLC

Dr. EKMAN. Thank you, Chairman Broun, Ranking Member Edwards. I really appreciate this opportunity to testify on this very important issue.

I have been working with TSA on SPOT for eight years based on 40 years of research on how demeanor—facial expression, gesture, voice, speech, gaze and posture—can help in identifying lies and also harmful intent. My research has examined four very different kinds of lies: lies to conceal a very strong emotion felt at that moment, lies claiming to hold a social political opinion the exact opposite of your truly strongly held opinion, lies denying that you have taken money that isn't yours, and lies in which members of extremist political groups attempt to block an opposing political group from receiving money.

Now, our research focuses on real-world lies that matter to society in which each person decided for him or herself whether to lie or tell the truth, just as we do in the real world. No scientist comes out of the clouds and tells us you are supposed to lie, you are supposed to tell the truth, except in experiments published in journals. The person who tells the truth knows that if he or she is mistakenly judged to be lying, they will receive the same punishment of the liar who is caught. This makes the truthful person apprehensive and harder to distinguish from the liar, just as it is in the real world. And the punishment threatened is as severe and highly credible to those who participate in the research as we could make it, passed by the University IRB.

I should mention I work in a medical school. I would never get it passed at Berkley, but at a medical school what I do is considered trivial.

Now, unlike any other research team, we have performed the most precise comprehensive measurements of face, gesture, voice, speech, and gaze, and those measurements have yielded between 80 and 90 percent identification of who is lying and who is telling the truth. The clues we have found are not specific to what the lie is about. As long as the stakes are very high, especially the threat of punishment, the behavioral clues to lying will be the same. It is this finding that suggested there would be no clues specific to

the terrorist hiding harmful intent than the money smuggler, the drug smuggler, or the wanted felon.

In my written testimony I raised three questions. First, what is the basis for the SPOT checklist? I have explained why I believe our findings on four very different kinds of lies provided a solid basis for reviewing what was on the SPOT checklist.

Question two, what is the evidence for the effectiveness of SPOT? Mr. Willis has already covered that. I won't attempt to repeat it. I am very eager to see that report that you are eager to see.

Question three, can SPOT be improved? That is a dangerous question to ask a scientist. We could always think that more research is necessary. But is it a wise investment compared to other things that the government can invest in regarding airport security? That is your decision, not mine. In my testimony I have outlined a couple of types of research that I think could be useful if you decide you would want to do more research. But we do not need to do more research now to feel confidence in this layer of security provided to the American people.

In my written testimony I attempted to answer questions that have been raised by critics of SPOT. Would it have not been better to base SPOT on how terrorists actually behave? Wasn't SPOT based on—Why wasn't SPOT based on people role-playing terrorists? Why is SPOT catching felons and smugglers, not just terrorists? And aren't people with Middle Eastern names or Middle Eastern appearance more likely to be identified by SPOT?

I would be glad in responding to questions to provide brief answers to each of these that are in my written testimony. Again, my thanks to the Committee and the staff of the Committee for the opportunity to talk to you and to the men and women in TSA who make flying a safer path than it would be without their dedicated efforts. Thank you.

[The prepared statement of Dr. Ekman follows:]

PREPARED STATEMENT OF DR. PAUL EKMAN, PROFESSOR EMERITUS OF PSYCHOLOGY, UNIVERSITY OF CALIFORNIA, SAN FRANCISCO, AND PRESIDENT AND FOUNDER, PAUL EKMAN GROUP, LLC

Thank you very much for the opportunity to provide information and testify before your subcommittee about the very important issues involved in TSA's SPOT Program. Here is an **Outline** of my written testimony:

1. Credentials.....page 1
2. My Research on Deception and Dangerous Intent.....page 4
3. What is the Basis for the SPOT Checklist.....page 7
4. What is the Evidence of the Effectiveness of SPOT.....page 9
5. Can SPOT be Improved.....page 9
6. Appendix.....page 10
   a. Comments on GAO report.....page 10
   b. Comments on NRC report.....page 10
   c. Comments on Jason report.....page 11
   d. Comments on Maria Hartwig's statements.....page 11
   e. Detecting lies in a counter-terrorism scenario: Body language. Frank, M.G., et. al.....page 12
   f. Deceiving about intentions in a security setting. Frank, M.G. et al.....page 15
7. References.....page 20

1 Credentials:

I received my PhD in psychology from Adelphi University in 1958, and after serving for two years as a 1st Lieutenant, U.S. Army Medical Service Corps, Chief Psychologist, Walson Army Hospital, I became part of the University of California Medical School, San Francisco (UCSF) in 1961. I retired from UCSF as a full professor in 2004.

My laboratory at UCSF was supported from 1961 through 1998, without interruption, by **grants** from the National Institute of Mental Health (NIMH), and at times by grants from NSF, and the Markle Foundation, and **contracts** from ARPA, DARPA, and DHS. The contracts from DARPA and DHS were specifically targeted on deception, the other contracts and grants supported basic research with direct application to deception.

I have received the following **honors:** identified by the American Psychological Association as one of the 100 most influential psychologists of the 20th century; Distinguished Scientific Contribution Award from American Psychological Association, 1991, (highest award for basic research); honorary doctoral degrees from the University of Chicago, 1994; University of Geneva, 2007; Adelphi University 2010; University of Lund, Sweden 2011.

My first **article on deception** [i] was published in a peer reviewed journal in 1969. Since then 15 articles on deception in which I am first or second author have been published in peer reviewed journals, and 13 chapters have been published in books on this topic. My **book** TELLING LIES[ii] was first published in 1985. It has never gone out of print in English, has been translated into more than a dozen languages, and is currently in a fourth edition (2009) with four new chapters not part of the first edition. My book WHY KIDS LIE[iii] was published in 1989, and has been translated into more than six languages.

Shortly after my retirement from University of California, San Francisco in 2004 I started a small company (Paul Ekman Group, PEG), which provides training -- through workshops and online tools -- on deception and demeanor and also on emotional skills. My goal was to translate the basic research studies I had conducted at UCSF into tools and workshops that could be of practical use. That intention is also manifest in the title of my book EMOTIONS REVEALED: Recognizing, Faces and Feelings to Improve

Communication and Emotional Life[iv] (2003); second edition with one new chapter in 2007.

I or my company has provided dozens of **workshops to law enforcement agencies** for thirty years, most recently (2010) to the New York Police Department and the Serious Organized Crime Agency (SOCA) in 2011, in London. I have provided workshops on **national security** to various agencies including CIA, FBI, MI-5 in London, and the Israeli National Police.

My focus in all this work is how **demeanor** – facial expression, gesture, posture, voice, gaze and speech – can provide clues to deception and dangerous intent.

While humans do not have Pinocchio's nose, there are signs that may be related to lying that always occur in everyone, what we call '*hot spots'*. These are signs in face, body, voice, speech, or the combination of these signs, that something is amiss, something of importance is happening, more than is being revealed. There are many reasons why hot spots occur, among them are lying about hostile intent. Thus the skilled observer who identifies a hot spot must then explore its nature to determine whether it is disguising some nefarious intention or whether it occurred for some other, non-harmful reason.

Currently my main focus is on the development on **online training** relevant to these topics. The Micro Expression Training Tool (METT)[v] and the Subtle Expression Training Tool (SETT)[vi] are currently available at my website and have been used successfully by tens of thousands of people worldwide. Research has shown that people can learn to spot concealed emotions from these online tools. Five new online training tools are currently under development by my company.

My association with SPOT began in 2003, initiated by an inquiry from Carl Maccario the person who originated the program. On a pro bono basis I observed passengers at Logan Airport and reviewed more than once and gave advice about the SPOT program. Again on a pro bono basis I have met with Behavior Detection Officers (BDOs) at various airports to hear their concerns and give them encouragement. The current contract to provide online training to TSA personnel, see Enclosure 3.

I have also consulted on the FAST programs, DHS's project on **automated physiological measurement** of malintent. If this program is successful I believe it will be a valuable adjunct but not a substitute for SPOT.

2 My Research on Deception and Demeanor

From the start of my research in 1967 it has differed from most other scientists studying deception and demeanor by focusing on **high stake lies**, in which the person lying has a lot to gain or lose by success or failure. Most other research on deception and demeanor have examined lies in which there is not much to lose or gain. My very first experiment took on the challenge of detecting a lie in which life itself was at stake – suicide. It was in my study of films of suicidal patients in the late 1960's that I uncovered the nature of **micro** facial expressions, very brief (1/25 second), expressions that leak concealed emotions.[vii] The research I designed studied the ability to conceal extremely unpleasant emotions, with the threatened punishment for failure -- the loss of professional career.

The next set of studies grew out of the consultations and training I was then providing to law enforcement and national security agencies in the late 1980s and 1990s. We specifically patterned the deception situations we employed after the types of criminal or intelligence gathering situations these agencies faced. For example, we gave volunteers the choice about whether to take or leave $50 in cash, and then lie or tell the truth about this theft, or we asked strong believers to lie or tell the truth about their strongly held opinions about a social issue (e.g. death penalty). This latter situation is comparable to the informant who tries to convince an intelligence officer of his true loyalties. In both scenarios, if they succeeded in deceiving the interviewer (who was me) they could earn $50, if they chose to tell the truth and the interviewer believed them they would earn $10. However, if the interviewer judged them to be lying, whether or not they were lying or telling the truth, they would receive no money, and they were threatened with severe punishments – locked in a totally dark room the size of telephone booth, subjected to 10-40 110 db blasts of white noise at random intervals - as loud as a firecracker, but just below the level that might cause hearing damage. Note that although we gave a sample of this punishment to each volunteer, we did not actually have to punish anyone. (I also note that this work was approved by the Institutional Research Board at my University complying with all federal guidelines about the ethical treatment of human subjects.)

This study and those that followed are unique in resembling the real world in three ways: (1) the research subject decides whether to lie or tell the truth. He or she is not ordered what to do by some authority figure (the experimenter). This is important because our early research suggested that different kinds of people choose to lie or be truthful. It is also important because it is a deliberately chosen act; in the real world, people choose to commit crimes or commit terrorism, they are not randomly assigned to

do so. (2) The punishment threatened is severe, realistic, and believed by the subjects, just as it is in criminal or terrorist situations. (3) Anyone judged to be lying faces punishment, regardless of whether the person actually lied or told the truth. As in real life, the innocent truthful person faces punishment if judged to be lying.

If these three features are not incorporated in a research study the findings are irrelevant to real world high stakes lies like those that SPOT is aimed to detect.

Our research program provided evidence very relevant to the **scientific underpinning of the SPOT** program. We found that the behavioral signs relevant to distinguishing lying and truthfulness are the same *regardless* of what the lie is about as long as there was a threat of severe punishment. The behavioral *hot spots* were the same regardless of whether the lie was about strongly felt unpleasant emotions, strongly held opinions or stealing money[viii]. This finding supported my prediction[ix] that when the stakes are very high, especially the threat of severe punishment if the lie is detected, it overloads a person's capacity to think clearly and regulate demeanor *no matter what the lie is about*. To repeat: we found that some basic core clues to deceit are not lie-specific but are the same across very different lies as long as there was a threat of severe punishment. Based on this evidence we expected the **terrorist** would show the same behavioral clues to deceit that we have identified in these other high stake lies.

The next study was specifically designed to provide information that would be most relevant to identifying terrorists. We [x] involved members of national security organizations in the U.S., England, and Israel in 2004 to advise us on designing research that would provide information *they* wanted to know. They all had personal experience dealing with terrorists; this included personnel from US Military intelligence, CIA, Scotland Yard, and Israeli National Police. .

In 2005, supported by DARPA, we (Professor Mark Frank, then at Rutgers, with myself as a consultant) recruited as research subjects members of **extremist** groups in the U.S. many of whom believe it is justified to break the law for their goals. They were given the opportunity to take a $100 check made out to a group that opposed them. If they took the money and succeeded in their lie, convincing an interrogator (usually retired FBI) that they did not take the money, the opposing group did not receive the money and their group did. In addition they personally received $75. They could also choose not to take the other group's money and if the interrogator believed they were telling the truth, the other group received $25, their group received $25, and they received $25. But if the interrogator disbelieved them, regardless of whether they were lying or being truthful, they were threatened with severe punishment and received no money.

Combining the measurements of face and body yielded a very high level of accuracy in identifying whether someone was lying or truthful; and replicated many of the same behaviors we identified in the earlier work. Although the findings are just now being submitted for publication, I have included excerpts from that publication (leaving out the academic and historical issues) as Appendix e.

The contract officer at DHS, who funded some of the analyses, required that the raw data from this study be given to the American Institute of Research – a nonprofit, independent research organization – so they could analyze and evaluate the methods and the data independently. They obtained the same very high level of accuracy in detecting lies from demeanor. It is rare – I know of no other example in any previously published behavioral study – when such independent scrutiny and verification of results is sought and obtained. Even though this work has not yet been submitted for peer review (but it will be shortly), I believe these findings should be regarded as solid.

I served only as a consultant on the next study carried out by Mark Frank now at the University of Buffalo, in which a situation closely resembling an **airport check point** was constructed, and people who belonged to political groups lied or told the truth about what they intended to do. (In the extremist study they lied or told the truth about what they had already done). As they waited in queue, a uniformed police officer passed by the queue and looked at each person. Frank analyzed the reactions of these people using some of the same behavioral measurement as in the extremist study, and found that these behavioral clues again distinguished the liars from the truth tellers, and overall at a slightly higher rate. Thus this evidence shows, as I predicted, that these behavioral markers can be useful even in situations where the person has yet to commit an illegal act. Those findings have not yet been submitted for publication, but a summary of the work is included in appendix f. We also note that the methodology of this study was also independently reviewed and approved by the American Institutes for Research.

I am currently developing a new test of the ability to catch a liar and an online training program to improve performance. Research to evaluate the impact of the training is planned. A second research project in development is to identify potential assassins in a crowd, and if that research is successful to then develop online training tools for alerting those doing surveillance to such persons.

3 What is the Basis for the SPOT Check List?

The check lists contains many of the behaviors which we have found in our studies of different types of lies – lies about emotions, strongly held opinions, taking money, and to deprive an opposing political group of income. Our research (see second full paragraph at the top of page 5) has shown that clues to deceit are **not lie-specific** but are the same regardless of what is being lied about, as long as the stakes are very high. Since these behavioral clues have been identified in multiple separate studies over a 30 year period, and since those in one of the studies were verified by an independent research group, to the extent that SPOT used our findings, we believe that part of the check list is on solid ground. All other behaviors listed on the checklist have been shown to differentiate liars and truth tellers in at least one published study, most of them by more than one. Rarely do applied materials in law enforcement settings contain as much scientific backing as this checklist.

I have been asked, would it not have been better to gather data on **how terrorists behave in airports**, and build the SPOT check list on that basis? Even if it was possible to mount surveillance cameras in every major airport in the U.S., it would have taken decades to accumulate enough behavioral records to analyze scientifically, since, fortunately, terrorism is a very rare event. For example, we know of 6 terrorists among the 29 million travelers who passed through Newark Airport in 2001. However, a place with more frequent terrorist concerns – Israel –were the creators of the behavioral observation system that eventually became SPOT. An Israeli who works in airport security has told my colleague that they based some of their system on my previously published work. And twelve years ago I taught Israeli security about my findings. The Israelis still use this system to date.

Some have wondered why research is not done to evaluate SPOT using people who **role play** being a terrorist, and see if they get through? The problem which renders that approach useless is that if the stakes are not very high, which they aren't in most role playing, the behaviors that betray a lie – many of them involuntary reactions - won't be generated. For example, my work 30 years ago showed that most people cannot raise the inner corners of their eyebrows on purpose. Yet, when people feel distressed, as liars often do, those movements will happen involuntarily. In the study of extremists (Appendix e.), there were 19 instances of this expression, and 16 of those 19 occurred in the liars.

I have also been asked: **why are felons and smugglers not terrorists being identified by SPOT?** The behavioral clues, or hot spots, are not specific to what the lie is about. A

basic set of core clues to deceit are the same regardless of what the lie is about if the threatened punishment is severe. The research evidence strongly suggests that there are no behavioral clues unique to terrorists that will not also be shown by a murderer, rapist, money smuggler, etc.

I was given the opportunity by English colleagues to view the surveillance videotapes of the London bombers taken shortly before they struck. Although the videotapes are of poor quality, what I was able to see suggested to me that SPOT personnel would have identified them. And the accounts from those who were at the feeder airport where the leaders of 9/11 boarded their flights to Logan airport, also suggested that they showed behaviors which would have been identified by SPOT if it had been in place at that time.

Some commentators on the SPOT program have claimed that those whose physical appearance and/or name suggests they might be from the Middle East might be apprehensive when entering an airport and therefore be more likely to be picked up by SPOT even though they are perfectly innocent. SPOT personnel are aware of this hazard. They know this is a behavioral profiling **not** a **racial profiling** program, and take account of the anxiety that might be felt by someone Middle Eastern in appearance. Also note that not all the behaviors on the SPOT list are anxiety based. Some years ago I suggested to the former director of TSA that research in airports should be done to insure that no racial profiling occurs. The idea was appreciated, the funding was lacking.

I have also been told by critics of SPOT that TSA should have first done observational research in airports, and the type of experimental check-point study carried out by Mark Frank and colleagues at Buffalo (on which I consulted; see page 6 and Appendix f.) before creating the SPOT program. That would be a great plan if Al Qaeda and associates agreed to a three year vacation, during which the American people would not need the layer of security provided by SPOT.

TSA was not groping in the dark when it initiated SPOT. It reached out for the best evidence available that would allow them to introduce this layer of security without delay. They came to me and my colleagues, based on their perusal of the scientific literature; I did not reach out to them to sell them anything. We were able to provide relevant information because our research showed that hot spots are useful clues that are not lie-specific but are present in all high stake lies when there is a threat of severe punishment. And finally, keep in mind that these behaviors do NOT trigger an arrest. They trigger a conversation, usually around 30-90 seconds in length, during which the Behavior Detection Officers attempt to ascertain why this individual showed the behaviors they did. At times they uncover malfeasance, at times they find an innocent

reason, at times they find a stressful but not illegal reason (e.g., a philandering traveler sneaking off to cheat on their husband or wife).

4 What is the evidence for the effectiveness of the SPOT program?

An extraordinarily impressive validation study was commissioned by Science & Technology of TSA, carried out by American Institute of Research, it is said that this report will be released April 1, 2011. I have not seen this report before submitting my testimony. And of course, deadlines for release of reports are not always met. I have been told about the report and I will describe below what I have been told.

In this huge study, 72,000 passengers who were selected at random (using an elaborate procedure that should have eliminated any bias in who was so selected), were compared to 23,000 passengers identified by SPOT. Malfeasants (felons, smugglers, etc.) were identified more than 50 times as often by those selected by SPOT. This finding provides **very important evidence for the validity of SPOT.** These findings also indirectly show that SPOT is alert to at least some of the right behaviors, for they would not have succeeded in this validity study if they were not doing so.*

The question should no longer be whether SPOT is effective – this report establishes that – but what can be done to make SPOT even more effective? In particular, are there any leaks in the system which can be identified – and then plugged – to provide even greater assurance that a terrorist will not get through.

5 Can SPOT be improved?

The answer is probably yes. Although my knowledge of what TSA is undertaking is by no means complete I do know that they are working on two very important issues: selection (how to identify for recruitment those most likely to perform SPOT best) and training (increasing substantially the amount of training provided to Behavior Detection Officers (BDOs). Establishment of a panel of expert advisors on how to improve the program is also underway.

---

* Critics have claimed that a terrorist was not identified at JFK, ignoring the fact that there were no SPOT personnel on duty at that time. Regrettably, there are not enough Behavior Detection Officers to observe all lines at all major airports.

There are many other steps that could be taken if there was the funding and the manpower. One study that especially interests me would reveal how often people who show many of the behaviors on the SPOT check list are **not** identified by the BDOs, essentially slipping through the net. If this occurs with any frequency, we need to know whether it is a function of the time of day, the number of hours a BDO has been working, the experience of the BDO, etc. Such a study would not demand very large resources, but this is only one of many research studies that could enhance SPOT, and investment in such research has to be balanced against other investments such as increases in training, increasing the number of BDOs, etc.

[I thank Professor Mark Frank for having critically reviewed my testimony and suggesting many useful additions and clarifications]

6 Appendix

There have been various reports and public statements criticizing the scientific basis for the SPOT program. I will briefly address some of them here.

a. GAO report

I was interviewed more than once by the authors of the report who I believe tried to provide a thorough evaluation of SPOT. However, I believe my views of SPOT as they emerged in the report were incomplete. Although my suggestions for further research were amply reported, my description of the evidence for the SPOT check list (see Section 3 of this report) were not adequately reported, creating the impression that I have serious doubts about the program and don't believe it is evidence based. I thought I made clear that in my judgment SPOT was the best that could be done given time urgency and financial constraints. Scientists enjoy spinning various new ideas for research, and I did that in my meetings with the GAO authors, perhaps unwittingly creating the impression that without that research SPOT was not on solid ground. Let me set the record straight. There is strong evidence, all of it published, some of it verified by other independent scientists, for the validity of the SPOT check list (Section 3 above); and, there is strong evidence that the SPOT program is effective (Section 4).

b. NRC report on the Polygraph

I was a member of the NRC panel, and I believe it is a superb evaluation of the validity of the polygraph in national security (there is no evidence of validity). The report much more briefly, and in a cursory fashion, considered

other approaches to detecting national security threats, including my work on demeanor. When that was considered I was absent due to prolonged illness. I believe the NRC report on deception and demeanor, the basis for SPOT, is not thorough, and the report writers did not have access to the information presented in sections 2 and 3 of this report.

c. Jasons Report

Although I have twice reported to the Jason's at separate meetings, I have not been given a copy of the 2008 report which is said to be critical of the science behind SPOT: "No scientific evidence exists to support the detection of inference of future behavior including intent". That quote, reportedly from a 2008 Jasons report, was in a 2010 *Newsweek* article. Note that the quote is about **future** behavior; there is a great deal of evidence about demeanor measures identifying lies about past behavior (section 2 and 3 of this report). At the time the Jason's report was written Mark Frank's study (described in section 2) had not yet been performed, which we now know did show success in predicting future behavior.

d. Maria Hartwig's criticisms

While Hartwig's own research has made some commendable improvements in research design on the issues of who can catch liars, and the strategies for doing so, she has dealt with low -not high- stake lies which have little relevance to my work or to the situation faced in SPOT.

In a 2011 TV interview Hartwig said: "The scientific research shows that it's very hard to detect whether somebody's up to no good just by looking at their behavior." She certainly is correct if the stakes are low; research by O'Sullivan, Frank, Hurley and Tiwanna [xi] has shown that when the stakes are low, law enforcement officers are not any better at detecting liars than laypeople. However, as I predicted, when the stakes are high these law enforcement officers clearly outperformed laypeople, likely due to the presence of many of these involuntary behaviors. Hartwig's research, as mentioned above, along with other deception research has usually dealt with low stakes lies and therefore likely did not elicit such behavioral clues. Jousting is not an academic sport I enjoy so I will go no further.

e. Detecting lies in a counter-terrorism scenario: Body Language

I have abridged this report, with Professor Frank's permission, excluding much important material, which is primarily relevant to an academic and/or scientific audience not a policy audience. Please contact mfrank83@buffalo.edu for the full report as it submitted for publication in an academic journal.

Frank, M.G.; Hurley, C.M., Kang, S., Svetieva, E., Sweet, D.M., Kim, D., Pazian, M. & Ekman, P.

Terrorism at its core is a human endeavor which can be successfully executed only if the terrorist escapes detection. This means a terrorist must successfully deceive a number of individuals along the way, including family, neighbors, local police, and security officers.

Counter terrorism scenario.

We derived the essential elements of our counter-terrorism situation based upon a two day meeting with working and retired counter-terror professionals from the USA, UK, and Israel, all of whom had previously spoken face to face with terrorists. We designed this scenario to match those sit down interviews they had each performed in the past. We identified a number of key characteristics that made this deception scenario novel in the research literature.

Group rewards and punishments. First, we recognized that religious terrorism, as that which occurred on September 11th, involves individuals who have different 'world views' than their opponents (REF). This means that the terrorist believes in the divine justice of his or her cause and group, whereas the opponents of the terrorist are not just infidels but directly defy the Supreme Being. This also means that the terrorist is concerned more about how his or her actions affect his or her group than how it affects him or herself. The way we imported these concepts into the laboratory was to recruit subjects who belonged to politically active groups, and told them that their interviewers were members of their arch rival group. So if the participant belonged to a pro-life group, then the interrogator belonged to, or was sympathetic to, a pro-choice group. This created the oppositional world view relationship. We then tailored the rewards and punishments to the group, and to a lesser extent, to the individual. The participants were asked to engage in an act hostile to the interests of their arch rival group – they were asked to steal a $100 check made out to their arch rival group, and then lie to an interrogator about the theft. If they were able to fool the interrogator, they were able to tear up the $100 check, and instead $100 would be directed to their own group. They would receive an additional $75 for their personal use. If they were not able to fool the interrogator, then they were told that the $100 would go to their arch rival group, and they would not receive any money, and they would be detained and face 30 minutes of loud startling blasts of noise. If they did not steal this $100 check, and were believed by the interrogator, they would receive $25 for their group, $25 would go to their archrival group, and they would receive

$25 for their personal use. If they did not steal this check, but they were disbelieved by the interrogator, they were told that their group would receive no money, they would receive no money, and they would face the noise blast punishment. They were also told that their arch rival group would receive an unspecified lesser amount of money. Thus we created a high stakes situation where the lie was designed to harm the oppositional group, help their own group, and subsequently themselves.

Choice. A lie is defined as a deliberate attempt to mislead, without prior notification (Ekman, 1985/2001). A liar who chooses to tell a lie, versus being assigned to tell a lie, would likely bear different feelings about the lie.

We predict that in this high stakes counter terrorism scenario, liars will show more **incongruous** behaviors – facial expression of emotions or symbolic gestures that do not fit with the words – than truth tellers. We consider any expressed emotion that does not accompany a statement referring to that emotion as incongruous. For example, the facial expression of fear is congruous with the statement "I was afraid of getting caught," whereas it is incongruous with the statement 'I just put the check back in the envelope.' We consider a symbolic gesture (like the A-OK finger to thumb) incongruous when the gesture means the opposite of the words spoken.

Method.

Participants. Our sample consisted of 132 participants (75 female and 57 male) who volunteered for a study entitled "Communication skills experiment." They all belonged to politically active groups who were identified on campus of a large public University in the Northeastern USA.

Procedure.

. The interrogators were male retired FBI or other law enforcement whose questions were scripted by the research team. We used these men because they spent their lives doing such interviews, and thus effected the behavioral disposition of a real law enforcement/terrorism interview.

Results.

Facial analyses. Table 1 shows the breakdown of participants who showed at least one or more negative facial expressions such as fear, distress, contempt, and disgust, which was incongruous with the words, by whether they chose to take the check or not the results were (75%; $X^2 (1) = 33.53, p < .001;$). We note that this accuracy based on the presence or absence of incongruous emotional expressions was just as high for truth tellers (79%) as it was for liars (72%). Thus, when answering a question in which a participant tells a confirmable lie, the presence or absence of a negative emotion can be very diagnostic of deception.

Symbolic gesture analysis. We only coded the yes and no gestures(head shakes and nods) as that allowed clear comparisons to affirmations and negations in the text. Table 2 shows that the pattern of incongruent gestures, and shows a significantly higher proportion of them when an individual was lying ($X^2 (1) = 10.47, p < .001$). We found that 78% of all incongruent symbolic gestures in this study occurred in the liars.

Combined analysis. Finally, we examined the interaction of the two types of incongruent expressions – facial expression of emotion and gesture to see how that affected classification accuracy.

We tabulated this by counting those participants show showed either a negative emotion, or an incongruent symbolic gesture , and compared them to those who showed neither. Table 3 shows that when we do that, accurately distinguish liars and truth tellers at 78%. When we add other measues we have of voice, speech and gaze we will be able to push this percentage up more than 10 points.

Table 1. *Presence or absence of fear, distress, contempt, or disgust that does not fit the spoken word by veracity.*

	Truth teller	Liar	Total
Negative emotion absent	48	20	68
Negative emotion present	13	51	64
	61	71	

X^2 *(1) = 33.53, p < .001 75% correct classifications*

Table 2. *Presence or absence of an inconsistent head gesture by veracity*

	Truth teller	Liar	Total
Inconsistent gesture absent	58	52	110
Inconsistent gesture present	3	18	21
	61	70	

X^2 *(1) = 10.47, p < .001*

f. Deceiving about intentions in a security setting.

I have again abridged this report with Professor Frank's permission, excluding details that are the same as in Appendix e., and also matters which are relevant primarily to an academic and/or scientific audience not a policy audience. Please contact markfrank83@buffalo.edu for the full report.

Frank, M.G., Hurley, C.M., Maroulis, A. Kim, D. [Paul Ekman served as a consultant]

This is the first experiment to test whether behavioral clues betray deception in a high stakes lie in which the individual has not yet not committed an act about which s/he will lie.

SUMMARY: In a situation set up to resemble an airport security context, we could predict at 90% accuracy who intended to lie about an action which s/he had not yet taken. This was accomplished by analysis solely on their emotional reaction, eye contact, and nervous body behaviors. These are the types of actions security officers look for in behavioral observation programs. These results are the first study to show that intentions can be detected from behavior.

The deception situation was patterned after an airport security checkpoint, where participants wait behind security lines, and then get approached by a security officer who passes by and looks at each of them. As explained in Appendix e, we once again recruited participants who were politically active, and again they had the opportunity, if they were successful in lying, to deprive an opposing group of receiving money and benefit their own group. (See Appendix e, page 19 for the details). Unlike the prior experiment, in this experiment they decided what they intended to do on the other side of the checkpoint. They had to pass through a stand up, check point screening, with the possibility of being either detained, sent to a secondary sit down interview, or allowed to pass through. After they had 10 minutes to think about the task, they would be asked to go into a room alone, and speak to a video camera outlining what they intended to do (i.e., to try to steal the check if they make it through, or to not try to steal it). This was patterned after the suicide bomber videos, which are used to help insure that the intentions are acted upon.

Method

Participants. The participants were 26 male and 15 females recruited from in and around the University at Buffalo. They all belonged to political advocacy groups,

Procedure. Participants and confederates (the participants did not know the other people in the lineup were employees of the research team) lined up, so that the participant was always second in the queue. They were told to wait until the security officer entered. The waiting area had a belt stanchion, and a metal detector that they would have to pass through before stepping up to a small podium to face the checkpoint officer. They waited approximately 10 minutes until the checkpoint officer emerged. These officers were active or retired police officers, all dressed in their uniforms. These officers emerged from a room approximately 5m in front of the queue, and the officer then walked toward the queue, approached

within 1-2m, and then looked at each of the 3 to 4 individuals in the queue as he walked down the line from the end to the beginning (thus looking at a confederate, a confederate, then the participant). The security officer then went behind the podium and called for all individuals in the queue to take out their ID's, and to remove any metal on their persons. An experimenter showed up then and collected the metal, as the confederates/participants walked through the metal detector. None of the security officers or members of the experimental team knew whether the participant had the intention of taking the check until after the entire experiment was over.

The segment that was analyzed for this project was the period of time marked by the opening of the door that exposed the presence of the security officer to the confederates and participant, until the point at which the security officer went behind the podium, This time period ranged from 15 to 44 seconds in length. There was no dialogue between the participants/confederates and the security officer.

Coding. The participants' behaviors in that 11-44 second critical period were coded by 2 expert FACS (Facial Action Coding System) coders who looked for Action Units associated with anger, contempt, disgust, fear, happiness, sadness, and surprise. They categorized the presence or absence of these emotions. They also coded the presence or absence of any self manipulative movements, including nervous hand wringing or tapping, and scored whether the participant made eye contact or averted his or her eyes, when the security officer looked at them.

Results.

There were no detectable sex differences, so we collapsed across that variable. When we summed across contempt, disgust, and fear, we found that negative emotions predicted who intended on taking the check, and who did not, at 85% accuracy; this pattern was statistically significant;(see Table 1.) Breaking that down further by emotion, we found that contempt/disgust only occurred in those intending on taking the check (8 out of 21), but never in the truth tellers (see Table 2). We found that fear reactions alone also occurred in 18 of the 21 participants who intended to take the check, and in only 3 of those with no intention (see Table 3).

We also examined nervous/fearful actions expressed in the body through hand wringing or tapping, manipulators, or gaze aversion. We found that 47.6% of the intention participants showed some hand wringing or tapping movements, whereas only 15% of the no intention participants showed these behaviors (see Table 5). We found that 38% of the intention participants avoided eye contact with the security officer on his pass through, whereas only 10% of the no intention participants avoided eye contact (See table 6).

Discussion.

The results showed clearly that one can detect the intention to steal a check from nonverbal behavior only. In a situation set up to resemble an airport security context, we could predict at 90% who intended to take the check, and who did not, based solely on their emotional reaction, eye contact, and nervous body behaviors. These are the types of actions security officers look for in behavioral observation programs. Thus, it is possible to show that intentions can be detected from behavior.

Table 1. Number of participants who showed contempt, disgust, or fear reactions by those who intended to take the check or not take the check.

	Intention	
	Take check	Not take check
Presence of negative emotion	18	3
Absence of negative emotion	3	17
Total participants	21	20

85.4% correct; (χ^2 (1) = 20.50, p < .001)

Table 2. Number of participants who showed contempt or disgust reactions by those who intended to take the check or not take the check.

	Intention	
	Take check	Not take check
Presence of contempt/disgust	8	0
Absence of contempt/disgust	13	20
Total participants	21	20

68.3% correct; (χ^2 (1) = 9.47, p < .003)

Table 3. Number of participants who showed fear reactions by those who intended to take the check or not take the check.

	Intention	
	Take check	Not take check
Presence of fear	18	3
Absence of fear	3	17
Total participants	21	20

85.4% correct; (χ^2 (1) = 20.50, p < .001)

Table 4. Number of participants who showed hand wringing or tapping reactions by those who intended to take the check or not take the check.

	Intention	
	Take check	Not take check
Presence of body actions	10	3
Absence of body actions	11	17
Total participants	21	20

65.9% correct; (χ^2 (1) = 5.03, p < .026)

Table 5. Number of participants who showed gaze aversion by those who intended to take the check or not take the check.

	Intention	
	Take check	Not take check
Gaze averted	8	2
Eye contact maintained	13	18
Total participants	21	20

63.4% correct; (χ^2 (1) = 4.39, $p < .04$)

7 References to Paul Ekman's Written Testimony

i. Ekman, P. & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88-105.

ii. Ekman, P. (1985). *Telling Lies: Clues to deceit in the marketplace, politics, and marriage*. New York: W. W. Norton. Latest edition 2009.

iii. Ekman, P. (1989). *Why kids lie*. New York: Scribners.

iv. Ekman, P. (2003) *Emotions Revealed: Recognizing faces and feelings to improve emotional life*. New York: Holt.

v. Ekman, P. (2004). *Micro Expression Training Tool*. (METT) http://www.paulekman.com.

vi. Ekman, P. (2006). *Subtle Expression Training Tool*. (SETT) http://www.paulekman.com.

vii. Haggard and Isaacs also reported seeing micro expressions but they considered them as signs of unconscious not deliberate concealment.

viii. Frank, M.G., & Ekman, P. (1997) The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology 72,* 1429-1439.

ix. Ekman, P. (1985). *Telling Lies: Clues to deceit in the marketplace, politics, and marriage*. New York: W. W. Norton. Latest edition 2009.

x. I and Mark Frank, a former post doctoral fellow who then took an academic post at Rutgers, worked jointly in designing this research and in planning data collection and analyses. Frank carried out the study, I advised on data analyses.

xi. O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior 33*(6), 542-543.

Chairman BROUN. Thank you, Doctor. I appreciate your testimony. I now recognize our next witness, Dr. Maria Hartwig, Associate Professor, Department of Psychology, John Jay College of Criminal Justice. Dr. Hartwig, your testimony for five minutes.

TESTIMONY OF MARIA HARTWIG, ASSOCIATE PROFESSOR, DEPARTMENT OF PSYCHOLOGY, JOHN JAY COLLEGE OF CRIMINAL JUSTICE

Dr. HARTWIG. Good morning. It is an honor to be here. Thank you for allowing me the opportunity.

The SPOT program is based on the idea that judgments of credibility can be made on the basis of observing facial cues and non-verbal cues that indicate stress, fear, or deception. And I have been asked to address the scientific support for this.

First of all, there are more than 30 years of research on deception that shows that people are quite poor at detecting deception on the basis of observing behavior. In a recent meta-analysis, a statistical overview of all the research, people obtained a hit-rate of 54 percent and you should, of course, keep in mind that 50 percent is the hit-rate you obtain by chance alone. So why are people so poor at detecting deception on the basis of observation? And one answer is that there are very few non-verbal demeanor-based cues to deception and these cues of deception tend to be weak. So simply put, there may not be much to observe. And contrary to what laypeople and presume lie experts such as law enforcement believe, liars don't display more signs of stress, fear, and arousal.

And critics of this research very often say that these findings are due to the nature of the laboratory experiments that most research relies on. And the claim is that when liars—when the stakes are sufficiently high, these cues to deception will appear. Research has addressed this concern by studying high-stake lies, such as lies told by people suspected of serious crimes like murder and rape, and these studies don't show any evidence that cues to stress and anxiety appear as the stakes increase.

And let me turn to the issue of detecting deception from facial cues to emotion. So this is based on the idea that liars experience emotion or fear of detection and that observing these facial cues can help you detect lies. I don't have time to go into details about the theoretical problems of that assumption, but in brief, it invites both missives and false alarms. It may miss travelers with hostile intentions who don't experience these emotions or who successfully conceals them and it may generate false alarms for travelers who don't have hostile intentions but experience these feelings for other reasons.

Most people are quite surprised to hear that there is very little evidence on the issue of these so-called micro-expressions, brief displays of an underlying emotion that are revealed automatically. I am aware of only one study published in the peer-reviewed literature conducted by Steve Porter and his colleague, Leanne ten Brinke, in the Journal of Psychological Science, they examined the prevalence of micro-expressions in falsified and genuine displays of emotion. They found no complete micro-expression in any of the 697 facial expressions they analyzed. They found 14 partial micro-expressions occurring in either the lower or the upper half of the

face, but these micro-expressions occurred with similar frequency in true and falsified expressions.

So this study shows that micro-expressions occur very rarely, and to the extent that they do occur, they occur in genuine displays as well. And the authors of this paper conclude that the occurrence of micro-expressions in true expressions makes their usefulness in airline security settings questionable. And they also state that the current training that relies heavily on the identification of full-faced micro-expressions may be misleading.

And finally, I would like to address a point of view expressed by Dr. Ekman in a recent article in Nature on the SPOT program. He stated that he no longer publishes all of the details of his work in the peer-reviewed literature because those papers are closely followed by scientists in countries such as Syria, Iran, and China, which the United States view as a potential threat. I object to deliberate strategy not to publish research for three reasons.

First, in that the enemy, whoever they are, a potential terrorist or criminals, may be aware of results from research applies to all deception research, so if we took this argument seriously, we shouldn't publish any lie-detection research because it may ultimately help the enemy.

And second, it is my understanding of the theory of micro-expression that these are automatic involuntary displays, and if that is the case, I fail to see how knowledge about these behaviors or the research on these behaviors could help the person.

And third and most importantly, these claims of micro-expressions as cues to deception or the cues included in the SPOT program, they are empirical questions that should be addressed with data and subjected to scientific peer review. And given the amount of resources that have already been spent on this program, I think such validation is absolutely necessary.

So in summary, my view is that the SPOT program is out of step with the scientific research. It relies on an outdated view of deception and there is very little support in the peer-reviewed literature. And if I had more time, I would say a few words about what I think may be a more productive approach to assessing credibility, but I believe I am out of time.

[The prepared statement of Dr. Hartwig follows:]

PREPARED STATEMENT OF DR. MARIA HARTWIG, ASSOCIATE PROFESSOR, DEPARTMENT OF PSYCHOLOGY, JOHN JAY COLLEGE OF CRIMINAL JUSTICE

The TSA has implemented the SPOT program, a security screening protocol that relies on observation of nonverbal and facial cues to assess the credibility of travelers. In particular, the program relies on behavioral indicators of "stress, fear, or deception" (GAO, p. 2). A key question is whether there is a scientifically validated basis for using behavior detection for counterterrorism purposes. This testimony will review the relevant empirical evidence on this question. In brief, the accumulated body of scientific work on behavioral cues to deception does not provide support for the premise of the SPOT program. The empirical support for the underpinnings of the program is weak at best, and the program suffers from theoretical flaws. Below, I will elaborate on the scientific findings of relevance for this issue.

Accuracy in deception judgments

For several decades, behavioral scientists have conducted empirical research on deception and its detection. There is now a considerable body of work in this field (Granhag & Strömwall, 2004; Vrij, 2008). This research focuses on three primary questions: First, how good are people at judging credibility? Second, are there be-

havioral differences between deceptive and truthful presentations? Third, how can people's ability to judge credibility be improved?

Most research on credibility judgments is experimental. An advantage of the experimental approach is that researchers may randomly assign participants to conditions, which provides internal validity (the ability to establish causal relationships between the variables, in this context between deception and a given behavioral indicator) and control of extraneous variables. Importantly, the experimental approach also allows for the unambiguous establishment of ground truth, that is, knowledge about whether the statements given by research participants are in fact truthful or deceptive. In this research, participants provide truthful or deliberately false statements, for example by purposefully distorting their attitudes, opinions, or events they have witnessed or participated in. The statements are subjected to various analyses including codings of verbal and nonverbal behavior. This allows for the mapping of objective cues to deception–behavioral characteristics that differ as a function of veracity. Also, the videotaped statements are typically shown to other participants serving as lie-catchers who are asked to make judgments about the veracity of the statements they have seen. Across hundreds of such studies, people average 54% correct judgments, when guessing would yield 50% correct. Meta-analyses (statistical summaries of the available research on a given topic) show that accuracy rates do not vary greatly from one setting to another (Bond & DePaulo, 2006) and that individuals barely differ from one another in the ability to detect deceit (Bond & DePaulo, 2008). Contrary to common expectations (Garrido, Masip, & Herrero, 2004), presumed lie experts such as police detectives and customs officers who routinely assess credibility in their professional life do not perform better than lay judges (Bond & DePaulo, 2006). In sum, that judging credibility is a near-chance enterprise is a robust finding emerging from decades of systematic research.

Cues to deception

Why are credibility judgments so prone to error? Research on behavioral differences between liars and truth tellers may provide an answer to this question. A meta-analysis covering 1,338 estimates of 158 behaviors showed that few behaviors are related to deception (DePaulo et al., 2003). The behaviors that do show a systematic covariation with deception are typically only weakly related to deceit. In other words, people may fail to detect deception because the behavioral signs of deception are faint.

Lie detection may fail for another reason: People report relying on invalid cues when attempting to detect deception. Both lay people and presumed lie experts, such as law enforcement personnel, report that gaze aversion, fidgeting, speech errors (e.g., stuttering), pauses and posture shifts indicate deception (Global Deception Research Team, 2005; Strömwall, Granhag, & Hartwig, 2004). These are cues to stress, nervousness and discomfort. However, meta-analyses of the deception literature show that these behaviors are not systematically related to deception. For example, in DePaulo et al. (2003), the effect size d (a statistical measure of the strength of association between two variables) of gaze aversion as a cue to deception across all studies is a non-significant 0.03. DePaulo et al. state: "It is notable that none of the measures of looking behavior supported the widespread belief that liars do not look their targets in the eye. The 32 independent estimates of eye contact produced a combined effect that was almost exactly zero (d = 0.01)" (p. 93). Moreover, fidgeting with object does not occur more frequently when lying, d = -0.12 (the negative value suggests that object fidgeting occurs less, not more frequently when lying, but this difference is not statistically significant), nor does self-fidgeting (d = -0.01) and facial fidgeting (d = 0.08). Speech disturbances are not related to deception (d = 0.00), nor are pauses (silent pauses d = 0.01; filled pauses d = 0.00; mixed pauses d = 0.03). Posture shifts are not systematically related to deception either, d = 0.05.

In sum, the literature shows that people perform poorly when attempting to detect deception. There are two primary reasons: First, there are few, if any, strong cues to deception. Second, people report relying on cues to stress, anxiety and nervousness, which are not indicative of deceit.

High-stake lies. Some aspects of the deception literature have been criticized on methodological grounds, in particular with regard to external validity (i.e., the generalizability of the findings to relevant non-laboratory settings, see Miller & Stiff, 1993) The most persistent criticism has concerned the issue of generalizing from low-stake situations to those in which the stakes are considerably higher. Critics have argued that when the deceit concerns serious matters, liars will experience stronger fear of detection, leading to cues to deception. There are several bodies of work of relevance for this concern. In a meta-analytic overview of the literature on credibility judgments (Bond & DePaulo, 2006), the evidence on the effects of stakes

was mixed: Within studies that manipulated motivation to succeed, lies were easier to tell from truths when there is relevant motivation. However, the effect size was fairly small (d = 0.17). However, when the comparison was made between studies that differed in stakes, no difference in lie detection accuracy was observed. Also, the meta-analysis revealed that as the stakes rise, both liars and truth tellers seem more deceptive to observers. That is, lie-catchers are more prone to make false positive errors - mistaking an innocent person for a liar - when judging highly motivated senders.

Furthermore, research on real-life high-stake lies, such as lies told by suspects of serious crimes during police interrogations, shows that people obtain at best moderate hit rates when judging such material (for a review of these studies, see Vrij, 2008). Behavioral analyses of the suspects in these studies do not support the assertion that cues to deception in the form of stress, arousal and emotions appear when senders are highly motivated. Vrij noted that the pattern from high-stake lies studies are "in direct contrast with the view of professional lie-catchers who overwhelmingly believe that liars in high-stake situations will display cues to nervousness, particularly gaze aversion and self-adaptors" (2008, p. 77). Moreover, he notes that the results "show no evidence for the occurrence of such cues" (2008, p. 77).

In sum, neither the research in general nor specific results on high-stake lies support the assumption that liars leak cues to stress and emotion, which can be used for the purposes of lie detection.

Verbal vs. nonverbal cues to deception

The SPOT program seems to rely heavily on evaluation of nonverbal cues. This emphasis on nonverbal behavior as opposed to verbal content cues runs counter to the recommendations from research. A number of findings suggest that reliance on nonverbal cues impairs lie detection accuracy. First, the meta-analysis on accuracy in deception judgments investigated accuracy under four conditions: a) watching videotapes without sound b) watching tapes with sound c) listening to audiotapes and d) reading transcripts (Bond & DePaulo, 2006). The accuracy rates in the first condition, where people based their judgments solely on nonverbal behavior, was significantly lower than in the other three, which did not differ significantly from each other. Thus, the combined results of hundreds of studies on lie detection suggest that having access to only nonverbal cues impairs lie detection accuracy.

Second, a number of studies have correlated lie-catchers' self-reported use of cues with lie detection accuracy. The purpose of such analyses is to investigate whether failure to detect deception coincides with the self-reported use of a particular set of cues. The results of these studies are consistent: They show that the more frequently a participant reports relying on nonverbal behavior, the less likely they are to be accurate in detecting deception. First, Mann et al. (2004) investigated police officers' ability to assess the veracity of suspects accused of murder, rape and arson. They found that successful lie detectors mentioned story cues (e.g., contradictions in the statement, vague responses) more frequently than poor lie detectors. Moreover, the more nonverbal cues the detectives mentioned (e.g., gaze aversion, movements, posture shifts), the lower their lie detection accuracy was. Second, Anderson et al. (1999) and Feeley and Young (2000) found that the more vocal cues lie-catchers mentioned, the more accurate they were in detecting deception. Third, Vrij and Mann's (2001) analysis of accuracy in judging the statement of a convicted murderer showed that the participants who mentioned cues to stress and discomfort obtained the lowest hit rates. Fourth, Porter et al. (2007) found that the more visual cues participants reported, the poorer they were at detecting deception.

It should be noted that reliance on nonverbal cues is associated not only with poorer lie detection accuracy, but also a more pronounced lie bias (a tendency to judge statements as lies rather than truths). That is, paying attention to visual cues increases the tendency for false positive errors - mistaking an innocent person for a deceptive one. This finding was obtained in one of the meta-analyses on deception judgments (Bond & DePaulo, 2006), as well as in a study of police officers' judgments of suspects of serious crimes (Mann et al., 2004).

The finding that reliance on nonverbal cues hampers lie detection is not surprising, given the research findings on cues to deception. These findings suggest that speech-related cues may be more diagnostic of deception than nonverbal cues (DePaulo et al., 2003; Sporer & Schwandt, 2006, 2007; Vrij, 2008). For example, DePaulo et al. (2003) showed that liars talk for a shorter time (d = -0.35), and include fewer details (d = -0.30). Liars' stories are also less logically structured (d = -0.25) and less plausible (d = -0.20). Liars and truth tellers differ in verbal and vocal immediacy (d = -0.55), and with respect to the inclusion of particular verbal elements, such as admissions of lack of memory (d = -0.42), spontaneous corrections

(d = -0.29) and related external associations (d = 0.35). These findings are in line with predictions from content analysis frameworks (e.g., Köhnken, 2004).

Detecting deceptions from facial displays of emotion

Theoretical concerns. Parts of the SPOT program seem to be predicated on the assumption that analyses of facial displays of emotion can improve deception detection accuracy. The claims of effectiveness for such approaches are not modest. In an interview with the New York Times, Ekman claimed that "his system of lie detection can be taught to anyone, with an accuracy rate of more than 95 percent" (Henig, 2006). However, no such finding has ever been reported in the peer-reviewed literature (Vrij et al., 2010). More broadly, there is no support for the assertion that training programs focusing on identifying facial displays of emotions can improve lie detection accuracy (Vrij, 2008).

Apart from lack of empirical support for the effectiveness of training programs focusing on the analysis of facial displays of emotion, there are theoretical problems with the approach. The assumption behind the training program is that concealed emotions may be revealed automatically, through brief displays sometimes referred to as microexpressions. Implicit in this assumption is the notion that liars will experience emotions, and that leakage of emotions can betray their deceit. This seems to equate cues to emotion with cues to deceit. But what is the evidence that lying will entail emotions, while truth telling will not? Several scholars have noted that the assumption that liars will experience emotion is a prescriptive view - it suggests how liars should feel. Common moral reasoning suggests that lying is "bad" (Backbier et al., 1997). In line with this reasoning, Bond and DePaulo (2006) proposed a double-standard hypothesis to explain the discrepancy between people's beliefs about deceptive behavior (that liars will display signs of discomfort and stress) and the actual findings on deceptive behavior (that liars typically do not display such signs). The double-standard hypothesis suggests that people have two views about lying: one about the lies they themselves tell, and one about the lies told by others (a form of fundamental attribution error; Ross, 1977). In the words of the authors: "As deceivers, people are pragmatic. They accommodate perceived needs by lying. [.] [Lies] are easy to rationalize. Yes, deception may demand construction of a convincing line and enactment of appropriate demeanor. Most strategic communications do. To the liar, there is nothing exceptional about lying" (p. 216). However, people's view of the lies told by others is markedly different: "Indignant at the prospect of being duped, people project onto the deceptive a host of morally fuelled emotions - anxiety, shame, and guilt. Drawing on this stereotype to assess others' veracity, people find that the stereotype seldom fits. In underestimating the liar's capacity for self-rationalization, judges' moralistic stereotype has the unintended effect of enabling successful deceit. Because deceptive torment resides primarily in the judge's imagination, many lies are mistaken for truths. When torment is perceived, it is often not a consequence of deception but of a speaker's motivation to be believed. High-stakes rarely make people feel guilty about lying; more often, they allow deceit to be easily rationalized. When motivation has an impact, it is on the speaker's fear of being disbelieved, and it matters little whether or not the highly motivated are lying (pp. 231-232)."

These are important points, in that they highlight the discrepancy between the perspective of the liar and the lie-catcher: People fall prey to an error of reasoning when assuming that the liars are plagued by emotions. They fail to take into account the pragmatic nature of lies, as well as the liar's ability to rationalize their lie. Moreover, they may misinterpret the fear of a motivated innocent person as a sign of deceit.

Beyond naïve moral reasoning about lies, is it psychologically sound to assume that people experience stress and negative emotion about lying? Can we expect that a criminal will experience guilt or shame about the actions he has committed, or that a prospective terrorist is plagued by negative feelings about the actions he is about to commit? They may, but given the double-standard hypothesis, we cannot be certain that this is the case. Apart from guilt and shame, it could be argued that liars may experience fear of not being able to convince. However, we must acknowledge the important fact that truth tellers might also experience such fear. For example, Ekman coined the term "Othello error" to describe how lie-catchers may misinterpret an innocent person's fear of not being believed as a sign of deception (Ekman, 2001). Moreover, people may react not only with fear but also anger in response to suspicion. Indeed, one study found that truth tellers reacted with more anger to suspicion than did liars (Hatz & Bourgeois, 2010). For an innocent person, suspicion is obviously undeserved. An emotional reaction to such treatment fits with a large body of social justice research suggesting that people have affective re-

sponses to violations of fairness (De Cremer & van den Bos, 2007; Mikula et al., 1998).

Empirical support. In sum, the concern raised above is that equating arousal, fear and stress with deception may rest on shaky theoretical grounds. If one rejects this concern and insists that such processes accompany lying, there is yet another hurdle to overcome. If people do experience affective processes, can they conceal them? Given the attention to microexpressions in the media, one might assume that there is an abundance of research published in peer-reviewed journals addressing this question. However, this is not the case. Porter and ten Brinke (2008) noted that "to [their] knowledge, no published empirical research has established the validity of microexpressions, let alone their frequency during falsification of emotion" (p. 509). They proceeded to conduct an analysis of people's ability to a) fabricate expressions of emotions they did not experience and b) conceal emotions that they did in fact experience. Their results showed that people are not perfectly capable of fabricating displays of emotions they do not experience: When people were asked to present a facial expression different from the emotion they were experiencing, there were some inconsistencies in these displays. However, the effect depended on the type of emotion people were trying to portray. People performed better at creating convincing displays of happiness compared to negative expressions. This is plausibly due to people's experience of creating false expressions of positive emotion in everyday life. With regard to concealing an emotion people did in fact experience, they performed better: There was no evidence of leakage of the felt emotion in these expressions. As for microexpression, no complete microexpression (lasting 1/5th-1/25th of a second) involving both the upper and lower half of the face was found in any of the 697 facial expressions analyzed in the study. However, 14 partial microexpressions were found, 7 in the upper and 7 in the lower half of the face. Interestingly, these partial microexpression occurred both during false and genuine facial expressions. That is, not only those who were falsifying or concealing emotions displayed these expressions; true displays of emotion involved microexpressions to the same extent. Porter and ten Brinke concluded that the "occurrence [of microexpressions] in genuine expressions makes their usefulness in airline-security settings questionable, given the implications of false-positive errors (i.e., potential human rights violations). Certainly, current training that relies heavily on the identification of full-face microexpressions may be misleading." (p. 513).

Passive vs. active lie detection

If it is difficult, or even impossible to detect deception through analyses of leakage of cues to affect, how can lie detection be accomplished? The research reviewed here suggests that it is more fruitful to focus on the content of a person's speech than to observe their nonverbal behavior, since the latter provides little valid information about deceit. The implication of this is that in order for lie judgments to be reasonably accurate, lie-catchers cannot simply observe targets. Instead, they should elicit verbal responses from these targets, as verbal messages may be the carriers of cues to deceit.

The proposition that lie-catchers ought to elicit verbal responses from targets fits with an important paradigm shift in the literature on deception detection. In brief, this paradigm shift involves moving from passive observation of behavior to the active elicitation of cues to deception (Vrij, Granhag, & Porter, 2010). This shift in the approach to lie detection is based on the now well-established finding that liars do not automatically leak behavioral cues. However, that the behavioral traces of deception are faint is not necessarily a universal fact: it may be possible to increase the behavioral differences between liars and truth tellers by exploiting some of the cognitive differences between the two. The approaches to elicit cues to deception are thus anchored in a cognitive rather than emotional model of deception. This model assumes that lying is a calculated, strategic enterprise that may demand cognitive and self-regulatory resources: Liars have to suppress the truth and formulate an alternative account that is sufficiently detailed to appear credible, while being mindful of the risk of contradicting particular details or one's own statement if one has to repeat it later on. Liars may experience greater self-regulatory busyness than truthful communicators, as a function of the efforts involved in deliberately creating a truthful impression (DePaulo et al., 2003).

Departing from this theoretical framework, it is possible to identify several different approaches to elicit behavioral differences between liars and truth tellers. First, if it is true that liars are operating under a heavier burden of cognitive load than truth tellers, imposing further cognitive load should hamper liars more than truth tellers. This hypothesis has been tested in several studies, in which cognitive load was manipulated (for example, by asking targets to tell the story in reverse order) and cues to deception were measured (e.g., Vrij et al., 2008; Vrij, Mann, Leal,

& Fisher, 2010). In support of the cognitive load framework, cues to deception were more pronounced, and veracity judgments were more correct in the increased cognitive load conditions.

A related line of research has investigated whether it is possible to elicit cues to deception by exploiting the strategies liars employ in order to convince. For example, this research has attempted to elicit cues to deception by asking unanticipated questions, based on the assumption that liars plan some, but not all of their responses (Vrij et al., 2009). In line with the predictions, liars and truth tellers did not differ with regard to anticipated questions, but when unanticipated questions were asked, cues to deception emerged. Moreover, liars' verbal strategies of avoidance can be exploited through strategic use of background information, which elicits inconsistencies or contradictions between the target's statement and the background information (Hartwig et al., 2005; 2006). For an extensive discussion on approaches to elicit cues to deception, see Vrij et al. (2010).

Summary and directions for future research

In summary, the research reviewed above suggests that lie detection based on observations of behavior is a difficult enterprise. Hundreds of studies show that people obtain hit rates just slightly above the level of chance. This can be explained by the scarcity of cues to deception, as well as the finding that people report relying on behavioral cues that have little diagnostic value. A wave of research conducted during the last decade suggests that lie judgments can be improved by the elicitation of cues to deception through various methods of strategic interviewing. This wave of research has been accompanied by a theoretical shift in the literature, moving from an emotional model of deception towards a cognitive view of deception.

The SPOT program's focus on passive observations of behavior and its emphasis on emotional cues is thus largely out of sync with the developments in the scientific field. The evidence that accurate judgments of credibility can be made on the basis of such observations is simply weak. Of course, it must be acknowledged that engaging travelers in verbal interaction (ranging from casual conversations to more or less structured interviews) is more time-consuming and effortful than simply observing behaviors from some distance. Still, the literature on elicitation of cues to deception suggests that this approach is likely to be substantially more effective than passive observations of behavior.

Evaluation of the SPOT program. At the time this testimony is written, the DHS's report on the validation of the SPOT program has yet to be released. Therefore, I cannot comment on the methodological merits of this validation study. However, as requested, I will briefly outline some methodological processes that I would expect a validation study to follow. First, it would be necessary to establish clear operational definitions of the target(s) of the program. What is the program supposed to accomplish? In order to evaluate the outcomes of the program, such definitions are crucial. Moverover, I would expect analyses of the outcomes of the SPOT program using the framework of decision theory. That is, a validation study should minimally provide information about the frequency of hits, false alarms, misses and correct rejections (to do this, one must have an operational definition of what a hit is). Those values should be compared to chance expectations based upon the baserate of the defined target condition. Then the obtained outcomes should be compared to a screening protocol that does not include the key elements of the SPOT program. For example, the outcome of a comparable sample of airports employing a random screening method may serve as an appropriate control group.

In addition to analyzing the results using a decision theory framework, it would be desirable to empirically examine the behavioral cues displayed by targets who pose threats to security, and compare them to targets who do not. That is, videotaped recordings of these targets (to the extent that they are available) should be subjected to detailed coding to determine the behavioral indicators that indicate deception and/or hostile intentions as these travelers move through an airport. The behaviors displayed by such targets should be compared to an appropriate control group, for example, a random sample of innocent travelers. The purpose of such analyses would be twofold: First, the results would empirically establish the behavioral indicators of deception and malicious intent in the airport setting. Second, the results could be compared to the SPOT criteria to establish whether there is an overlap between the two sets of indicators.

Moreover, it would be useful to evaluate the criteria on which Behavior Detection Officers rely to make judgments that a target is worthy of further scrutiny. That is, analyses of the behaviors of targets selected for scrutiny could be subjected to coding, to establish a) whether the officers rely on valid indicators of deception and hostile intentions and b) whether they rely on the criteria set forth in the SPOT training program. This would validate the SPOT program in a slightly different

manner, as it would assess to what extent the Behavior Detection Officers follow the protocol of their training.

A problem of using field data is that important data will likely be missing. That is, while databases may include information about hits and false alarms from travelers who are subjected to further scrutiny, the data on misses and correct rejections are will be incomplete. For example, misses may not be detected for years, if ever. For this reason it may be appropriate to subject the SPOT program to an experimental test, in which the ground truth about the travelers' status is known. The field and experimental approaches are obviously not mutually exclusive: It is possible (and perhaps even preferable) to conduct both types of validation studies, as the strength and weaknesses of each approach in terms of internal and external validity complement each other. A multi-methodological approach to validating the SPOT program may also provide convergent validity. If a concern with the laboratory approach is that participants in an experimental study would not be sufficiently motivated, it may be worth mentioning that it is possible to experimentally examine the effect of motivation on targets' behaviors within the context of a laboratory paradigm. Some targets could be randomly assigned to receive a weaker incentive for successfully passing through the screening, while others receive a stronger incentive. Of course, it would not be possible to create a fully realistic incentive system due to ethical considerations. Still, such a manipulation could provide some insight into the role of motivation in targets' behaviors, and to what extent motivation moderates the display of relevant behavioral cues.

In closing, I will briefly note a few areas of relevance for the airport security screening settings that I believe future research ought to focus on. First, most research has examined truths and lies about past actions. In the airport setting, truths and lies about future actions (intentions) may be of particular relevance. A few recent studies have examined true and false statements about future actions (Granhag & Knieps, in press; Vrij, Granhag, Mann, & Leal, in press; Vrij et al., in press). The studies reveal some findings in line with the research on true and false statements about past actions, for example in that false statements about intentions are less plausible (Vrij et al., in press). However, there are also some differences in these results. While research on statements about past actions shows that lies are less detailed than truths, this finding has not been replicated for statements about future actions. However, this body of work is still small, and further empirical attention is needed. Second, and relatedly, it would be valuable to attempt to extend the research findings on elicitation of cues to deception to airport settings. That is, it would be useful to establish to what extent it is possible to increase cues to deception using cognitive models when the statements concern future actions. Such knowledge could be translated into brief, standardized questioning protocols that could be used to establish the veracity of travelers' reports about both their past actions and their intentions.

References

Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom? Journal of Nonverbal Behavior, 23, 67-89.

Backbier, E., Hoogstraten, J., & Meerum Terwogt-Kouweenhove, K. (1997). Situational determinants of the acceptability of telling lies. Journal of Applied Social Psychology, 27, 1048-1062.

Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. Personality and Social Psychology Review, 10, 214-234.

Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. Psychological Bulletin, 134, 477-492.

De Cremer, D., & van den Bos, K. (2007). Justice and feelings: Toward a new era in justice research. Social Justice Research, 20, 1-9.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. Psychological Bulletin, 129, 74-118.

Ekman, P. (2001). Telling lies: Clues to deceit in the marketplace, politics and marriage. New York: Norton.

Feeley, T. H., & Young, M. J. (2000). The effects of cognitive capacity on beliefs about deceptive communication. Communication Quarterly, 48, 101-119.

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgments: Accuracy and estimated ability. International Journal of Psychology, 39, 254-275.

The Global Deception Research Team (2006). A world of lies. Journal of Cross-Cultural Psychology, 37, 60-74.

Government Accountability Office (2010). Aviation security. GAO-1-763.

Granhag, P. A., & Knieps, M. (in press). Episodic future thought: Illuminating the trademarks of true and false intent. Applied Cognitive Psychology.

Granhag, P. A., & Strömwall, L. A. (2004). The detection of deception in forensic contexts. New York, NY: Cambridge University Press.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. Law and Human Behavior, 30, 603-619.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Deception detection via strategic disclosure of evidence. Law and Human Behavior, 29, 469-484.

Hatz, J. L., & Bourgeois, M. J. (2010). Anger as a cue to truthfulness. Journal of Experimental Social Psychology, 46, 680-683.

Henig, R. M. (2006). Looking for the lie. New York Times, Feb 5.

Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth'. In P.A. Granhag, & L.A. Strömwall (Eds.), The detection of deception in forensic contexts (pp. 41-63). Cambridge: Cambridge University Press.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. Journal of Applied Psychology, 89, 137-149.

Mikula, G., Scherer, K. R., & Athenstaedt, U. (1998). The role of injustice in the elicitation of differential emotional reactions. Personality and Social Psychology Bulletin, 24, 769-783.

Miller, G. R., & Stiff, J. B. (1993). Deceptive communication. Newbury Park: Sage Publications.

Porter, S., & ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. Psychological Science, 19, 508-514.

Porter, S., Woodworth, M., McCabe, S., & Peace, K. A. (2007). "Genius is 1% inspiration and 99% perspiration".or is it? An investigation of the impact of motivation and feedback on deception detection. Legal and Criminological Psychology, 12, 297-310.

Ross, L. D. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 10), pp. 174-221. New York: Academic Press.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-analytic synthesis. Applied Cognitive Psychology, 20, 421-446.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. Psychology, Public Policy, and Law, 13, 1-34.

Strömwall, L. A., Granhag, P. A., & Hartwig, M. (2004). Practitioners' beliefs about deception. In P. A. Granhag & L. A. Strömwall (Eds.), The detection of deception in forensic contexts (pp. 229-250). New York, NY: Cambridge University Press.

Vrij, A. (2008). Detecting lies and deceit: Pitfalls and opportunities (2nd ed.). New York, NY: John Wiley & Sons.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (in press). Lying about flying: The first experiment to detect false intent. Psychology, Crime & Law.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. Psychological Science in the Public Interest, 11, 89-121.

Vrij, A., Leal, S., Granhag, P. A., Fisher, R. P., Sperry, K., Hillman, J., & Mann, S. (2009). Outsmarting the liars: The benefit of asking unanticipated questions. Law and Human Behavior, 33, 159-166.

Vrij, A., Leal, S., Mann, S., & Granhag, P. A. (in press). A comparison between lying about intentions and past activities: Verbal cues and detection accuracy. Applied Cognitive Psychology.

Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. Applied Cognitive Psychology, 15, 187-203.

Vrij, A., Mann, S., Leal, S., & Fisher, R. P. (2007). "Look into my eyes": Can an instruction to maintain eye contact facilitate lie detection? Psychology, Crime & Law, 16, 327-348.

Vrij, A., Mann, S., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. Law and Human Behavior, 32, 253-265.

Chairman BROUN. Thank you, Dr. Hartwig. If you want to add some suggestions, we would be glad to enter those in the record and entertain those suggestions that you may have. And hopefully, we can get those from you.

Now, I would like to recognize our final witness and that is Dr. Philip Rubin, Chief Executive Officer of Haskins Laboratories. Dr. Rubin, you have five minutes for your oral testimony.

TESTIMONY OF PHILIP RUBIN, CHIEF EXECUTIVE OFFICER, HASKINS LABORATORIES

Dr. RUBIN. Chairman Broun, Ranking Member Edwards, and distinguished Members of the Subcommittee, thank you for the opportunity to speak to you today. My name is Philip Rubin. 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 activities previously mentioned by Chairman Broun, I am also a member of the Technical Advisory Committee that was formed to provide critical input related to analyses and methodologies used in the SPOT program.

I was invited here today to describe the current state of research in science and the behavior 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 written testimony provides some brief historical background on selected activities in the behavioral sciences related to security and it mentions a variety of documents and reports, some of which I have here, include many produced by the National Academies National Research Council, such as consensus reports and other documents. But the written testimony focuses on two that I was involved with: a workshop on field evaluation in the intelligence and counterintelligence context, and a short set of papers on threatening communications and behavior. Because of time limitations, I am not able to describe these in detail and refer you to my written testimony.

Regarding the field evaluation workshop summary, however, a number of the participants spoke about various obstacles to field evaluation, 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. Perhaps the most basic obstacle is simply a lack of appreciation among many for the value of objective field evaluations and how inaccurate informal "lessons learned" approaches can be to field evaluation.

A number of people throughout the process of developing this summary spoke about the pressures to use new devices and techniques once they have become available because lives are at stake.

This sense of urgency can lead to pressure to use available tools before they are evaluated, and it can even lead to ignoring the results of evaluations if they disagree with the user's conviction that the tools are useful.

As indicated earlier, I am a member of the Technical Advisory Committee for SPOT. As the GAO report indicates, the Technical Advisory Committee's role is extremely limited. It focused in the main on determining whether or not the research program successfully accomplished the goal of evaluating whether SPOT can identify high-risk travelers—defined as individuals who are knowingly and intentionally attempting to defeat the airport security process. The advisory committee has not been asked to evaluate the overall SPOT program, nor has it been asked to evaluate the validity of indicators used in the program, not asked to evaluate consistency across measurement, field conditions, training issues, scientific foundations of the program, and/or behavioral detective methodologies, et cetera. In order to appropriately scientifically evaluate a program like SPOT, all of these and more would be needed.

To summarize my written testimony, I would like to just mention a few points as highlights. These are some recommendations of how to move forward, so I am just going to hit some bullets.

First, 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 result as open as possible is necessary for determining if these technologies and devices are performing in a known and reliable manner.

Incorporate knowledge on the complexities, subtleties, irregularities, and idiosyncrasies of human behavior.

Next, understand the interplay and differences between affect, emotion, stress, and other factors.

Make sure that we are not distracted or misled by the tools and toys that fascinate us.

Pay serious attention to the ethical issues and regulations related to human subjects research, including 45 C.F.R. 46, the Common Rule, where applicable, and relevant emerging areas, including privacy concerns, neuro-ethics, and ethical implications of the deployment of autonomous agents and devices.

Reduce conflicts of interest to the extent possible, including financial conflicts of interest.

Develop an understanding of how urgency, organizational structure, and institutional barriers can shape program development and assessment.

And support the importance of the need for independent evaluation of new and controversial projects and issues with appropriate scientific, technical, statistical, and methodological expertise.

Thank you.

[The prepared statement of Dr. Rubin follows:]

PREPARED STATEMENT OF DR. PHILIP RUBIN
CHIEF EXECUTIVE OFFICER, HASKINS LABORATORIES

Chairman Broun, Ranking Member Edwards, and Members of the Subcommittee on Investigations and Oversight of the Committee on Science, Space, and Tech-

nology, 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 on issues that affect all of our citizens' lives. Often this advice takes that form of published documents known as consensus reports. 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 reports 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 and Threatening Communications reports.*

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 report summarized the meeting and highlighted key issues. Following [single-spaced sections] are extracts/adaptations of the *Field Evaluation* Workshop Summary, edited for continuity [attribution quotes 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 atDefence 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 area 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 sub optimal heuristics, and over reliance 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 be stand 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 AssessmentScreening 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 micro tremors-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 micro tremors are suppressed. Thus by monitoring the micro tremors 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 micro tremors in the voice, then they must result from tremors in some part of the vocal tract-the larynx, perhaps, or the supra laryngeal 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 micro tremors 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 micro tremors in the vocal tract, which would produce micro tremors in the voice. However, research has found no such micro tremors, 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 micro tremors 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 carryout 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 to the 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 DefenseCounterintelligence 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 the9/11 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 textanalysis 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 Bioethics*, 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, Divi-

sion ofBehavioral 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 EmergentNeurophysiological 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 andEducation. National Academies Press, Washington, DC.

National Research Council. (2008). *Human Behavior in Military Contexts.* Committee on Opportunities in Basic Research in the Behavioral and SocialSciences for the U.S. Military. Board on Behavioral, Cognitive, and SensorySciences, 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 ofInformation 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 andEducation. 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 PublicFigures.* National Academies Press, Washington, DC.

Russell, James A., Bachorowski, Jo-Anne, and Ferńandez-Dols, Jośe-Miguel. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology,* 54, 329349.

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.

Chairman BROUN. Thank you, Dr. Rubin. And I want to express my appreciation for your being here. I know you have had some recent challenges and I greatly appreciate you being here in spite of those. So thank you very much.

Dr. RUBIN. Thank you.

Chairman BROUN. And I want to thank all the panel for your testimony. Reminding Members that the Committee rules limit questioning to five minutes. The Chair at this point will open the round of questions and the Chair recognizes himself for five minutes.

Mr. Willis, when can we expect the SPOT validation report?

Mr. WILLIS. The report was delivered to me by AIR last night. It is being submitted through DHS's review and release distribution process. I am not exactly sure what that time is or when it is ultimately disseminated. I can certainly get that information for you, sir.

Chairman BROUN. I would appreciate getting that report to us as quickly as possible.

Mr. WILLIS. Yes, sir.

Chairman BROUN. What additional steps have to be taken before we get the report?

Mr. WILLIS. I don't know what DHS's distribution process entails. I know that I will submitting it this morning following my participation here.

Chairman BROUN. Do you have any problems in releasing the preliminary results?

Mr. WILLIS. I don't know what DHS's policy is on that, but I am happy to provide whatever is consistent with DHS's S&T's policy on release.

Chairman BROUN. I understand that the results, I assume, are still preliminary. There appears to be a discrepancy in the SPOT's success rate. In your testimony you state "the study did indicate that a high-risk traveler is nine times more likely to be identified using Operational SPOT versus random screening." Yet when you met with the staff from the I&O Subcommittee on March 3 you said that the SPOT program was 50 times more effective than random screening. One of our other witnesses, Dr. Ekman, also makes a similar claim in his testimony saying "malfeasance, felons, smug-

glers, et cetera, identified more than 50 times as often by those selected by SPOT." Can you please explain the discrepancy?

Mr. WILLIS. Well, there shouldn't be a discrepancy. We use four metrics by which to evaluate SPOT. The first one was the possession of illegal or prohibited items. The second one was possession of fraudulent documents. The third was LEO arrest, law enforcement arrest. And the fourth was a combination thereof. The LEO arrest has the higher number that you referred to in your question, sir.

Chairman BROUN. The 50 times?

Mr. WILLIS. Yes, sir. The possession of prohibited items and fraudulent documents is approximately four and a half times, and if one combines all of them, it is nine times.

Chairman BROUN. Are those that were identified—how many of those were actually convicted?

Mr. WILLIS. Sir, I would have no idea. Our effort stops at whether a decision is recorded as being arrested or not, and that is the information that is available through the SPOT database. It doesn't go beyond that.

Chairman BROUN. Do you have any data about false negatives? I mean false positives?

Mr. WILLIS. On?

Chairman BROUN. On the people that have been identified at the 50 times or 9 times or 4–1/2 times?

Mr. WILLIS. Are you talking about the false positive associated with arrests?

Chairman BROUN. No, with arrest or—yes, sir, with arrest and with prosecution—the ultimate prosecution, et cetera.

Mr. WILLIS. Yes, sir. We do have information available on that. So for example, if one looks at the false positive index, which is for every person that you correctly classify as a high-risk traveler, what is the number of travelers you misclassify? We have that information on any of the four metrics that we discussed. And so for example, combined outcome for every person that you correctly identify using Operational SPOT, 86 were misidentified. For the base rate or random study, for every person that you correctly identify, 794 were misidentified.

Chairman BROUN. Wow. SPOT was initially developed as intended to stop terrorism. That is the whole point of it. Now, we see that the program has expanded to include criminal activity. Why was this done?

Mr. WILLIS. You are asking a question about the mission. I am from Science and Technology, sir. I am unable to answer that. May I refer you to TSA?

Chairman BROUN. Well, that is the reason TSA should be here and the reason that I think Ms. Edwards and I are both extremely disappointed that they are not here.

Mr. WILLIS. I could, sir, talk to you about why we use metrics that deal more with criminal than with terrorism.

Chairman BROUN. That would be sufficient—or helpful.

Mr. WILLIS. Sure.

Chairman BROUN. You have got a few seconds, so go ahead.

Mr. WILLIS. Okay, sir.

Chairman BROUN. My time is out.

Mr. WILLIS. The reason we use those metrics that we had just listed, sir, was because they were available to us through the data in sufficient numbers to analyze, even though they themselves are low base rate or extremely rare. And data directly dealing with terrorism is unavailable and, thus, can't be used as a metric.

Chairman BROUN. Okay. My time is up. Ms. Edwards.

Ms. EDWARDS. Thank you, Mr. Chairman. And as I mentioned earlier, I am disappointed that TSA isn't here because I think that there are a number of questions that actually go to things like training protocols and other aspects of the SPOT program that they would have, you know, really useful information to share and so I look forward to working with the Chairman and the Committee.

This question about who needs to appear or not is not a decision, really, for the Administration. Congress determines, under its Constitutional authority, who appears before the Committees and what the jurisdiction is. So I do share that concern.

I want to go to this question, though, of profiling——

Chairman BROUN. Does the gentlelady yield?

Ms. EDWARDS. Yes.

Chairman BROUN. I appreciate your comment. You took up about almost a minute with that and I would like to give you an extra minute on top of that, so I don't want to charge you that time.

Ms. EDWARDS. I appreciate that, Mr. Chairman.

Chairman BROUN. So I will give you the extra minute. So if you all would start her clock again, please.

Ms. EDWARDS. Thank you. Thank you again, Mr. Chairman. I have a question, really, that goes to this issue of profiling. I mean, as an African American woman who sometimes, because I have short hair and I get cold, I wear a scarf on my head and that is true in the airports especially. I have had the experience of actually being pulled over, questioned, and it hasn't just happened once or twice. It has actually happened multiple times. And, you know, I don't want to make any speculation about that, but it does raise the question of who is identifying me and how and what I am sending off.

I am also reminded in Dr. Hartwig's testimony that, you know, I remember when I broke a lamp and I tried to glue it together and my mother walked in and she said what did you do? And I suspect that part of the reason that she could say that and she knew—and then I proceeded to tell her a lie, but I suspect that part of the reason that she knew I was lying is because she knew me and because she had had experience with me and because she had read my both verbal and nonverbal cues many times over, which gave her a much better indication of when I was doing truth-telling and when I wasn't.

We don't have that experience in our airports, and so I have a question for Lieutenant DiDomenica, and that is whether it is possible to train officers of all kinds not to engage in profiling? And I have done police training, law enforcement training as well, and I think it is tough to train out culture, culture in the sense of a police culture and a law enforcement culture where you have to train against type when it comes to these issues. And so I am curious, Lieutenant DiDomenica, if you can share with us whether it is possible to train officers not to engage in profiling?

Mr. DIDOMENICA. I believe it is so and I have been training in biased policing and racial profiling for over a decade now. Principally, with the state police I designed statewide programs for the Massachusetts police community on racial profiling, biased policing, and it is possible to make people aware of their own unconscious bias and tendency to want to make snap decisions about people based on very superficial things. We all have this hardware, it is a survival instinct, and when we look at somebody, we are automatically making an opinion about them. And a lot of it has to do with our background and cultural influence, and a lot of those are negative. But, you know, this part of your brain is about survival, and it wants to understand what is going on very quickly. And it actually gets a jump on your conscious awareness. So right away when I walked in here and you saw me and I saw you, we made a decision about each other before we were even consciously aware of who we were and what we are. And that is going on all the time. And this is the source of bias.

Now, knowing that I can't stop my feelings about someone based on how they look, that initial survival reaction about whether the person might be dangerous or not, but I can take a few seconds, maybe minutes, to think about, you know, what is going on, what do I know objectively, and maybe even do some race transposition. If this person was another race, you know, how would I feel about the situation? And then I can make a decision. So it takes self-awareness. It takes training. It takes the ability—willing to change and monitor yourself. But it can be done.

One of the foundations of the behavior assessment training I have done and what I initially gave to TSA for the SPOT program is you have to address bias and racial profiling. In fact, I call it—you know, it was—to me it was an antidote to racial profiling——

Ms. EDWARDS. Lieutenant DiDomenica, I would love to hear but I just have just a minute and a half left and I wanted to get to—I appreciate your answer. I wanted to get to Dr. Ekman because I have to tell you, you have been unnerving me the entire time I have been in here and I am sure we have been reading those cues. And I wonder if you have something to share with us on this issue of whether you can train against those kind of—what could be negative instincts in one context but train them to be positive factors in recognizing behavior?

Dr. EKMAN. Yes. And thanks for the opportunity to respond to that. I wanted to quickly put in that we did research years ago that show that the better you knew someone, the worse you were in identifying when they lied to you because you are biased. If they are your friend, your spouse, et cetera, you don't want to discover that they are lying. Strangers do better than close people.

But the issue is monitoring—building into the SPOT program some monitoring to discover the actual incidents of racial profiling. And my bet is that some people show a lot more of it than others. Not everybody can learn everything. Not everybody can unlearn everything. What we want as BDOs are the people who have the flexibility of mind to benefit from that training and be susceptible to racial profiling. How can we find out? It is not rocket science. It is by having unannounced observers checking on who is it they pay attention to and finding out whether there are some people

who are repeatedly showing racial profiling. And you either reeducate or you reassign them to a different job.

Ms. EDWARDS. Thank you, Dr. Ekman, and thanks for your indulgence, Mr. Chairman.

Chairman BROUN. You know, we will always be friends and I will always give you some variances on the time so I am not going to be worried about that at all.

Dr. Benishek, you are up next for your questions. Go ahead, sir.

Mr. BENISHEK. Thank you, Mr. Chairman. Thanks to the panel, as well, for being here.

It is our job here to try to spend the money of the taxpayer the most efficacious way and listening to the testimony here, it is really difficult for me to determine whether this SPOT process is accurate or not. But I would like to address Mr. DiDomenica about the process a little bit more. From your comments today it seems as if there is some doubt, I mean, even after the BDO sees some kind of behavior, then what is the process after that? If there is someone there, it sounds as if you have some doubt as to the next step as to what is happening, the next screening step. Are those people not trained in the same thing? I mean I would hate to see somebody get missed. So I would like to know more about the exact process from the moment that the person gets taken out of the queue. Is that effective? Is it—are we doing any good? Are we missing people? I mean, this is the kind of thing I think you brought up in your testimony.

Mr. DIDOMENICA. I think it is effective and I also think we are missing people, but I think that could be improved. The process actually starts with an observation that may indicate a person that is high-risk, that maybe should not get on that airplane or get onto that train or into that government building, whatever the critical infrastructure is. And based on the evaluation, this SPOT scoring, which I really can't go into because that is, you know, that is sensitive information.

But there are two levels, and one is more screening, and one is a law enforcement response. So for the people deemed to be the most high-risk, the protocol is to invite or call a law enforcement officer to do a follow-up interview. Now, this follow-up interview is the opportunity to address the false positives, because a lot of people that exhibit the behaviors that may indicate possible terrorist intent or criminal intent are just people that are upset or distracted or late for work or going to a funeral, whatever it is, that maybe a lot of people just get on the radar. And this interview, which really only takes a couple of minutes to do, is the opportunity to resolve that so you are not creating false positives. And it is also an opportunity to determine if you have got the real thing, that this person is high-risk. And so that is another skill. I mean that is the interview skill, which is another part of this process. So there are——

Mr. BENISHEK. Are those people skilled enough in your opinion?

Mr. DIDOMENICA. When you say "those people"——

Mr. BENISHEK. The people—the secondary person. Are there enough of those people?

Mr. DIDOMENICA. I think the responsibility ultimately falls on police officers when there is a high-risk person. I think they are ca-

pable. Every day they are making decisions around this country whether to arrest somebody, not to arrest somebody, use lethal force in some cases, deny people their freedoms, and so I don't think it is too much to ask them to make a decision, is this person a high-risk person and do we need to slow down the process to figure out what is going on? I think they are capable of doing it. We are doing it—whether this program gets funded or not, cops are making these decisions every day. But I would like to see them get more training and more support to make them better at what they do. And this program has that potential.

Mr. BENISHEK. All right. Thank you. I don't know where we are at with the time, but I will yield back the remainder of my time, if any.

Chairman BROUN. Thank you, Doctor. I just want to say your questioning just shows further why TSA should be here so that we could answer those questions, because if they were, then you could direct it to the TSA individuals and it would be very instructive to the whole Committee, Democrats and Republicans alike, and help us to go forward.

The next person on the agenda is my friend, Mr. McNerney. You are recognized for five minutes.

Mr. MCNERNEY. Thank you. And I appreciate you calling this hearing. It is interesting. I have watched "Lie to Me" on occasion and I find it is compelling but not too scientific in my opinion. But it is good for us to examine this issue and see how much utility there can be from it and how much money should be expended to find that utility.

Dr. Hartwig, I think I heard you say—and you can correct me if I am wrong—that you fail to see how knowledge of the indicators could be useful.

Dr. HARTWIG. I think that is, again, an empirical question. There isn't enough research on—well, there is a lot of research on demeanor cues, but as far as I know, there is no study that tests whether knowledge about, for example, micro-expressions help people not display them. But that would be a second step. It would be a good first step to establish that these expressions occur reliably.

Mr. MCNERNEY. Okay, and I was——

Dr. HARTWIG. So countermeasures come second.

Mr. MCNERNEY. Okay. Thank you, Dr. Hartwig. And I was going to follow up with you, Dr. Ekman, to basically say would you agree that knowledge of those indicators would also be useful to potential wrongdoers?

Dr. EKMAN. We don't know. I mean you are basically asking the question in polygraph terms is could you develop countermeasures?

Mr. MCNERNEY. Right. Right.

Dr. EKMAN. A proposal I put in to the government to find out—I mean I have reason to believe that the Chinese know the answer because they were sending me questions that you would want to prepare on if you were going to do a training study to see whether you could inhibit people from showing not just micro-expressions but there are dozens of items on that checklist. The—our government has not decided that it is worth finding out whether you can beat the system. Other governments are finding out and may be selecting people who can and training them so they can. We just

don't know. We know about the polygraph. We know countermeasures are quite successful. We know about some verbal means and we know they are quite successful.

If I can have a moment more, sir.

Mr. MCNERNEY. Yeah, go ahead.

Dr. EKMAN. You heard some complete contradictions between Dr. Hartwig and myself. I think if you look carefully at the literature, you would find that it comes out supporting me. But how can you know? And I think you need to do, when you get a disagreement among scientists, is you need to establish an advisory panel, experts, who have no vested interest and no connections to hear from the people who disagree and look at the literature and resolve it because you are really being given, in this testimony, advice that is 180 degrees opposite in terms of is there a scientific basis for what is being done?

But you could argue—and I don't know whether Mr. Willis would—that if this validity study holds up to scientific scrutiny, to everyone who has looked at it, to this Committee, if it is as successful as the report is, you have got to be doing something right to get that kind of success. So maybe it——

Mr. MCNERNEY. It——

Dr. EKMAN. —is of scientific interest to find out.

Mr. MCNERNEY. Thank you, Dr. Ekman. Mr. Lord is chomping at the bit here. Go ahead.

Mr. LORD. I would like to respond to Dr. Ekman's point. In fact, that was the key recommendation of our May 2010 report was to have an independent panel review the results of this current AIR validation effort. We think it is very important for a panel to be established that has no ties to the current program, that is not an advocate of the current program, to help weigh in on this very issue. I think it is very interesting that the panel today shows a lack of consensus, which was the basic point I made in my earlier statement. There is no scientific consensus——

Mr. MCNERNEY. Well, a subject like this you would expect to be—a broad range of disagreements. Has a panel—like what you are recommending—been suggested in one of the budgets or lined out somewhere or is this something——

Mr. LORD. Yeah, DHS agreed to establish an independent panel to review the methodology of the AIR validation effort, as well as to review the final results, but as Mr. Willis indicated, the final results of this latest validation effort have only recently been submitted. I believe he said as of last night.

Mr. MCNERNEY. I think I have run out of time so I am going to yield back.

Chairman BROUN. Mr. Hultgren, five minutes.

Mr. HULTGREN. Thank you, Doctor. Thank you all for being here. I share the frustration with some of the others that TSA is not here today. I am a new Member here at Congress, along with quite a few others, and so have been traveling much more in the last 3 months than I have ever traveled in my life. In fact, just on Monday, the trip out here, I had my first experience of the full treatment by TSA out of O'Hare and it was interesting. Didn't realize that it involved turning your head and coughing, but I now know that that does—is what it is. But, you know, it is important for us

to have these discussions again to protect our liberty and freedom, while at the same time making sure that we have security. So I do thank you for your role. What I am learning is that we have got a lot more work to do and a lot more discussion that needs to take place.

I just have a couple questions. Dr. Rubin, if I can address my questions to you if that would be all right. Much has been made about the science and research behind the ability for an individual—or in this case, BDO—to detect emotion, deceit, and intent in another individual based on a combination of verbal and nonverbal and micro-facial expressions. I wonder, speaking broadly and keeping it as simple as you can for those of us laymen, could you just tell us the state of the science as it relates to the detection of emotion, deceit, and intent of behavioral cues?

Dr. RUBIN. Yes. In general I guess I would agree with Dr. Ekman in the sense that we are at the point where there are two things going on. If you look at something like voice stress analysis and look at the meta-analysis done by Sujeeta Bhatt and Susan Brandon coming out of the Defense Department. What you basically see in most of these studies is that the results are no different than chance. Agreeing with both Dr. Hartwig and Ekman, there is a lot of controversy here and there is very little real science and validation.

And it is not just that field evaluation when you can't do it. Again, there has been a committee established on the SPOT Program regarding the report. I am on that committee. And we have not been asked to do any overall scientific validation for the program, just to look at one particular thing, are the results different than chance? So I am agreeing here that what is really needed on these issues, before we continue to invest more money, is to really establish, without putting any information at risk, a baseline about what is doable, what is not doable, what is known, and what is not.

So this is the classic issue of do you test first and then field a product or project? Or field it and test? And this particular instance, considering the investment, considering the intrusion on people's privacy, I think it is absolutely time to be testing, validating, and scientifically exploring these things now before we continue to do significant investment. I am not saying we shouldn't continue the program. I think it is important. But right now we need to establish on some of the known kind of things that we are doing without giving anything away. Is there good science behind it? Otherwise, we are simply throwing money down the drain.

Mr. HULTGREN. I think kind of following up on that, one of the concerns that operators have is that behavioral science is not dismissed because there are issues dealing with the validation of specific cues. Can you speak for a moment on the importance of behavioral science in counterterrorism context and then what its limitations are, what its strengths are as far as our work for counterterrorism?

Dr. RUBIN. Okay. Well, we are changing the topic a little bit because we are moving to counterterrorism. I think that the behavioral work is broad in counterterrorism. I think it is extremely important. Again, when we get to counterterrorism, you are broadening your argument out because you get to analysts. There has

been an excellent report from an NRC Committee chaired by Barouche Fish. There is a lot that is known.

And again, we touched on some of this and a number of the panelists did. You are starting to get involved in behavioral issues of attitude, of biases. Some of this was described in the original intelligence work of Richards Heuer on cognitive biases. There is a lot that we know. The issue becomes structural and organizational.

Consider, two things. What do we know? And what don't we know? With the stuff that we do know, how do we make sure it is being most effectively used by the intelligence community and by whomever else needs to use it on those issues where we are not entirely clear? Where things are uncertain or controversy, how can we move ahead? And then there are emerging technologies that we are going to start to be seeing used. We see some of them in terms of the kind of devices like x-ray, but things like euro-imaging, remote imaging, and sensing of other things. That is where I was speaking of the seduction of technology. I support that stuff greatly, but we need to make sure on stuff that is new and emerging that we also get a handle on it.

So I think the behavioral tools and technologies are stuff is growing rapidly, and are extremely important, but I think we are not developing a comprehensive approach to appropriately evaluating them before deploying them in the field.

Mr. HULTGREN. I see my time is up. I do want to thank you all for being here. I do feel like this is a start of a discussion that we need to continue, so I appreciate so much all of you being here. I also would ask for any advice any micro-facial expressions I might have so I don't have to go through that examination again. That would be helpful. So pass that along to me. Thank you.

Chairman BROUN. Thank you, Mr. Hultgren. I ask unanimous consent that the gentleman from Florida, Mr. Mica, be allowed to sit on the dais with the Committee and participate in the hearing. Hearing none, so ordered. Mr. Mica, you are recognized for five minutes.

Mr. MICA. Well, thank you. And first of all, thank you, Mr. Chairman, Mr. Broun, and Ranking Member Edwards and other Members of the panel.

I have great interest in the subject that you have before you. As you may know, I was involved in the creation of TSA when I chaired the Aviation Subcommittee in 2001 for some six years after that and watched its evolution.

First, I might say that I am absolutely distraught that your Subcommittee would be denied by TSA the opportunity for them to be here and possibly learn something or participate. I don't want you to feel like they are just ignoring you. They have ignored our Committee and others, so they have a history of this. And I will work with you and others. In fact, I think we need to convene a panel of Chairs of various Committees and somehow rein this Agency in. And it has an important mission. I am just stunned, again, that they would not have someone at least to hear from the excellent panel of witnesses you have had here today, particularly when they come and ask for more money.

Let me just tell you my involvement with the SPOT program, again, as Chair of the Committee that created it. I followed TSA

in its successes and failures and we have deployed a lot of expensive technology out there, and unfortunately, the technology does not do a very good job and the personnel failure performance rate is just off the charts.

And if you haven't had the classified briefing on the latest technology, which are both the backscatter and the millimeter wave, I urge you to do that. I had GAO review that in December of last year and then the pat-down, which was sort of their backup new procedure, which they put in place the end of last year. And then I had that reviewed by GAO in January. But that failure rate is totally unacceptable.

The way we got started on SPOT is I found the technology lacking in reports of performance both by screeners and the equipment they used as leaving us vulnerable, particularly after the Henchmen bombers. And I think we bought some puffer machines at the time. I remember going up, having those tested. They didn't work but they promised me they would. They deployed them and they didn't work. So we needed something in place. We encouraged looking at the Israeli model and you can't really adopt the Israeli model because they have a much smaller amount of traffic. We have 2/3 to 3/4 of all the passenger traffic in the world and that is part of America. You know, you get on a plane, you go where you want. People just have a magic carpet through aviation in this country.

That is how we started this. I have observed their operations and I can't evaluate them. We had GAO evaluate them and you have some representatives here to tell you that the failure rate is unacceptable. It is almost a total failure. If it wasn't money and personnel, maybe it wouldn't matter, but they have got 3,300 SPOT officers, I believe, in the program and they have got a quarter of a billion dollars in expenditures and asking for more.

What I heard today is that, again, it doesn't work. I had to leave before I heard all the suggestions and I would look for—. Some of the suggestions on the amount of time to do a verbal interview would improve it, but maybe finding some way to get us to a number that we could have some exchange.

Ms. Edwards made some excellent points in her opening comments, too, that we have got to have some way to improve this and that unless there is some verbal exchange, I think that we are with this standoff observation, we are wasting time, money, and resources. So I don't have a specific recommendation for the replacement. I do know what is in place does not work. But I can't tell you how much I appreciate your Subcommittee taking time to review this matter and try to seek a better approach, a better science, and better application of something that is so important. Because we are at risk. These people are determined to take us out.

I just came from another meeting, the folks that developed both backscatter and millimeter wave, which is two technologies we are using, and the scary thing there is we had witnesses in one of the other hearings that said that both of those technologies will not be able to detect either body cavity or surgical implants. And we already see that they are always going one step ahead of whatever we put in place. So we have got a failed system, we are spending a lot of money on it, it is supposed to provide us with a backup.

The information we have and the review of the performance shows that it is not doing that and it needs to be replaced or dramatically revised if it is going to be effective in keeping us from this next set of threats.

So those are my comments. I would ask that if you have suggestions, we do have an FAA bill which we can include some positive suggestions. We couldn't do that in the House side because of jurisdiction, but we can do it in conference and the door has already been opened by the Senate. And I would love to hear recommendations from you and from those who participated today how we can do it better. So thank you for allowing me to participate.

Chairman BROUN. Well, thank you, Chairman Mica. I appreciate your being here and appreciate your comments. I can speak for Ms. Edwards. We both are very concerned about national security. We both are concerned about civil liberties. We both are concerned about that we make sure that the flying public are safe and I appreciate her input. And I hope that you will find some way that maybe we will have those terrorists subjects that we can put in a study so that maybe some kind of behavioral science could be developed to try to identify these folks.

We will go to our next round of questioning. So I will recognize myself for five minutes for questioning. Even if SPOT is more than nine times more effective than random, we still are talking about very low base rates. Lieutenant DiDomenica who states in his testimony that the base rate for terrorism is .000000—I think one more 0, 6—I hope I didn't get too many zeros and did not leave that one. Can any of the panelists help put that into perspective? Anybody? Mr. Lord?

Mr. LORD. Sure. That statistic implies that acts of terrorism are very rare events. That makes it very difficult to test the efficacy of the program and develop, as we recommended in our report, performance metrics to allow you to better judge whether the program works as designed. But we don't think that should deter you from trying to craft what we would call proxy measures, other measures that help you get at this at least indirectly. And we made that very important recommendation, and TSA and DHS agreed to try to develop these indicators.

There is one step we think they could take that would make this exercise a lot more useful, currently they use a very long list of behaviors, the exact number and the characteristics are considered sensitive security information. But we posed a question, how do you know this is the right number? And they also assign point scores to each of these behaviors. Again, the details are sensitive security information. But that would be one way that we think would make the program more useful in identifying potential acts of terrorism, validate the point system, scrub the list of behaviors, cull the list, and try to come up with something that is more related to an eventual arrest or a hostile act. And there are ways to do that statistically.

Chairman BROUN. Thank you, Mr. Lord. Anybody else? Mr. Willis, yes?

Mr. WILLIS. Thank you, Mr. Chairman. So first off, proxy measures are a standard part of research, especially in the area of terrorism, because again, there are no direct measures in sufficient

quantities, typically, to use for terrorism. Criminal activity is often used as a proxy measure. It is an accepted practice mainly because when one is looking for terrorism or acts of terrorism in a lot of transit areas, you are looking for somebody who is coming in to try to use some false identification or you are looking for somebody who is smuggling. And both of these things are represented in higher numbers, even though they are still low base rate numbers in criminal activity. And so that is why that is typically used and used by other organizations as proxy measures. So I want to make sure that we were comfortable that we had given forethought to that and used what is a best practice for proxy measures, sir.

Chairman BROUN. Dr. Ekman?

Dr. EKMAN. There are a number of organizations. I work with airport security in England. I have seen the videos of the bombers before they bombed. I have worked in Israel where they do a lot of, of course, security. But even within our own government, the different parts of DOD that deal with counterterrorism and the attempts to identify terrorists in field military situations, there is no sharing of information. There is a lot of information out there that hasn't been brought together. It is sensitive, but it needs to be brought together and then with that database, take a look at what is on the SPOT list. I haven't seen what is on the SPOT list for four years so I don't know how it has changed and I don't know how it has been informed by research findings from our group and other groups and from observations by Special Forces, by our counterintelligence, by NYPD. There is a lot of information in this country in separate little pockets that hasn't been brought together.

Chairman BROUN. Thank you. My time has expired. For my questioning now, I recognize the Ranking Member, Ms. Edwards, for five minutes.

Ms. EDWARDS. Thank you, Mr. Chairman. I want to go to a question that was raised by Mr. Mica's comments when he was here. And I just want to be clear that from the perspective of GAO and the report and analysis that you have done, Mr. Lord, we don't yet know if the SPOT program is "a fiasco." Isn't that correct?

Mr. LORD. Yes, that is absolutely correct. Those were his words. That is not in our vocabulary. Thank you.

Ms. EDWARDS. And just to be clear again, what metrics again would you use to determine the success or failure as an operational program?

Mr. LORD.Since we have identified several instances of terrorists transiting through the U.S. system, studied the videotapes of their movement. Are they, in fact, exhibiting signs of stress? Are they, as some literature suggests, they don't typically emote much because they believe they are going on to a more blissful state. So it is unclear to us at this juncture whether there would be discernible signs of stress or fear. But there is videotape evidence that would allow you to get at that and we think that would be invaluable in fine-tuning the program.

Ms. EDWARDS. Yeah, I think I highlighted that in your testimony because there are a number of examples that we have. And I wonder, Mr. Willis, has DHS made an attempt to pull together not just video evidence here in the United States but with our international partners to do some kind of an assessment stacked up against the

screening techniques that have been identified to see whether we are on target? It is an awful lot of money to spend without, you know, putting it up against real-time data.

Mr. WILLIS. Thank you. Again, I represent DHS Science and Technology, not the operational community. From a——

Ms. EDWARDS. This is a science question.

Mr. WILLIS. Yeah, from a Science and Technology perspective, we are attempting to locate video of terrorist threats in other countries, as well as within the United States. And it is very difficult to try to get access to that information or to successfully get access to that video. And so if——

Ms. EDWARDS. Well, part of the reason that we pulled DHS together is because it was—you know, because it is a, you know, a collection of all of our, you know, sort of security and investigative interests under one house to work with our international partners. And so it is a little staggering to me to know that you have not had the capacity in now a decade to look at video and use it to make an analysis about whether the techniques that you seem to be employing are—would be successful. I mean that seems to me kind of a basic scientific question that DHS should be in a position with our partners internationally and here in the United States to get that video and, you know, conduct some real scientific analysis of that. So I would urge DHS to consider that.

I want to go to Dr. Hartwig for a minute because in your testimony you indicated that there are some other recommendations that you might make and I wonder if you could just describe very briefly those to us because I don't think you had an opportunity here in your testimony.

Dr. HARTWIG. Right. I think it is roughly captured by what Mr. Mica said before he left, that is it important to engage a person in conversation to elicit cues to deception. Overall, the research shows that statements carry some cues to deception. And also there is an emerging wave of new research that focuses on how to create cues to deception, how to elicit cues to deception because there is such an abundance of research showing that people don't just automatically leak. So my basic answer is that some form of questioning protocol, some kind of brief interview protocol that is based on the scientific research on how to elicit cues to deception, how to ask questions so that the liars and truth-tellers respond differently. I think that would be a worthwhile enterprise.

Ms. EDWARDS. So you are not really saying—and this is a yes and no—saying scrap the program, but you are saying that there are areas where we need to significantly improve the techniques that we are using to take us down a track of really being able to identify potential terrorists?

Dr. HARTWIG. Yes, I think if efforts would be spent on the questioning part of the program, that would put it much more in line with the scientific research.

Ms. EDWARDS. Thank you. Thank you, Mr. Chairman.

Chairman BROUN. Thank you, Ms. Edwards. We have been joined by the Congresswoman from Florida, Ms. Adams. You are recognized for five minutes.

Mrs. ADAMS. Thank you, Mr. Chair. Mr. Willis, earlier you said that there had been 71,000 referrals and you made a distinction of

that, the behavior leading to arrest. How many of those were arrested?

Mr. WILLIS. Of the 71,000?

Mrs. ADAMS. Yes.

Mr. WILLIS. That is the random selection method.

Mrs. ADAMS. Correct.

Mr. WILLIS. 71,000 were referred in the random selection. Nine arrests were made.

Mrs. ADAMS. Nine?

Mr. WILLIS. Yes.

Mrs. ADAMS. And in the other method?

Mr. WILLIS. Using SPOT 23,000 and a little bit were referred and 151 were arrested.

Mrs. ADAMS. And the types of arrests?

Mr. WILLIS. I don't have the nature of the arrests in the data that we looked at, ma'am.

Mrs. ADAMS. So it could have been belligerency or any other thing for that matter?

Mr. WILLIS. Some of them were for prohibited items that were on them at the time. Others could have been through outstanding warrants or something of that nature, ma'am.

Mrs. ADAMS. Do you think that I have an appearance or would I be a target for SPOT? I mean every time I go through the airport I get pulled aside and searched. And the reason I ask that is because, you know, being a past law enforcement officer and trained, I have some concerns about the way you are identifying pulling people aside. Dr. Hartwig, you said you wanted—you thought the program would work if more tools were available. Would it be better to use a validated system as opposed to one that is untested and invalidated?

Dr. HARTWIG. Well, first of all, I didn't say that about that the program would work. I was talking about where I think more emphasis should be spent or put.

Mrs. ADAMS. So even with the more emphasis do you believe that it would work?

Dr. HARTWIG. I don't know. I think we would need a properly conducted study to find that out. And I think it would be important to go beyond examining the arrest rates and to look at what are the actual behaviors that are displayed by these people who are arrested and to compare those behaviors with those that are in the list of queues. I don't know what those queues are because it is not available. And to look at are the SPOT criteria actual indicators. So I think that—it is definitely—we need to know whether it works or now.

Mrs. ADAMS. Mr. DiDomenica, you are a law enforcement officer. I am a past law enforcement officer. Do you believe that the TSA employees have enough training and the skills sets based on the training they are receiving to—you know, to provide this type of screening at this level?

Mr. DIDOMENICA. I think with a proper follow up by trained law enforcement that they do. But if we don't have the proper follow up by the police officers to figure out what is going on because this is just like an alarm. It is like going through the magnetometer and beeps. Well, what does that mean? So someone comes over and

pats you down. Well, the cops are like the pat-downs. All right. Why did this beep? And so if you have that level of follow up by trained law enforcement, I am comfortable with the training they receive. But without that level of follow up, I am not comfortable.

Mrs. ADAMS. So would it be your opinion that there needs to be more training?

Mr. DIDOMENICA. Yes.

Mrs. ADAMS. I yield back.

Chairman BROUN. Thank you, Ms. Adams. Mr. Willis, I have got another question for you. Does TSA plan to use R and D to improve the SPOT program or does it believe the program cannot be improved upon?

Mr. WILLIS. We do have some ongoing research with them and if I may say this is one of the beginning research elements that we have with TSA, sir, and in fact it was started in 2007 prior to GAO's interests. Its focus is specific, not to evaluate absolutely everything going on with SPOT. That is a huge tasking of which we are not tasked or resourced to do. This is looking at the indicators, the checklist itself, the existing checklist.

The first question that needs to be asked from a scientific perspective is does the checklist as it is currently put together and as it is currently deployed accomplish its mission. You would like to be able to compare that against random and against something else that has been shown to be out there and valid, but the fact is that there isn't another behavioral-based screening out there employed by any other group that we are aware of, either in the United States or abroad, that has been statistically validated. And so we have not been able to address that. So we compared this against random, which is the first scientific basis.

Chairman BROUN. So TSA is doing research?

Mr. WILLIS. We are doing research that supports TSA.

Chairman BROUN. Ms. Edwards, do you have another question?

Ms. EDWARDS. I do, thank you, Mr. Chairman. I just want to follow up with you, Mr. Willis, because I am confused. My understanding is that you shared with our staff that there is a pool video available of suicide bombers and the like that could be used to study. And I mean I would expect that if TSA were operating the right kind of way that would also be used for training. And so I am a little confused by your answer and I just want to be clear. Do we have video both from ourselves and perhaps from our international partners that we could use to assess the techniques that have been developed and the questions that—the assessment questions that have been developed so that we can make sure that we have a program that is working as effectively as we know it can work?

Mr. WILLIS. We don't presently have a sufficient number of videos to conduct scientific analysis on. S&T is attempting to work with our partners in the United States and internationally to gather these, but being a resource organization, we do not have the ability to compel operational organizations, much less international ones to provide us with that video. What we are doing is attempting to continue to collect that at—the best we can, as well as to conduct other kinds of supporting things such as interviews of direct eyewitnesses to suicide bombings, international subject matter

experts in the area to go beyond what the current validation study was, which is of the existing indicators, to try to help establish from a scientific perspective what is being used operationally abroad and, in fact, what is being witnessed by, again, eyewitnesses and subject matter experts so that we may be able to then bring that information back and test it to see——

Ms. EDWARDS. Is S&T doing that or TSA? Who——

Mr. WILLIS. That is S&T research, ma'am.

Ms. EDWARDS. Okay. And so I guess I mean for the—for our Drs. Hartwig and Ekman, it would be useful, wouldn't it, to have a pool, a real data pool to be able to assess that and develop a research protocol that enabled us to stack our assessment tools against that? And so my question, though, for Mr. Willis whether or not—what agency do you think is—would be the responsible one to get this pool together? Is it DHS? Is it TSA? Mr. Lord?

Mr. WILLIS. I don't know the right organization for that.

Mr. LORD. In our report, we made 11 recommendations. One of the recommendations was to use and study available video recording to help refine the SPOT program. In their formal Agency comments, the Department indicated they agreed and they were taking steps to do that so I think the Department is already on record for saying they agreed. It is a good idea. We are going to do it. So I mean they are—they bought into this idea. To the extent they have actually implanted it, we will have to follow up and see the extent they have addressed it. But just so—to clarify, DHS has bought into this idea. They have already agreed to do it.

Ms. EDWARDS. Thank you. And then finally, Mr. Lord, since you already have the microphone, DHS hasn't done a cost/benefit analysis on the program or a risk assessment. And it is my understanding that they don't do a great job actually—and I apologize for the critique—of either conducting cost/benefit analyses or risk assessments for many of their programs. How do we know if we even need the program?

Mr. LORD. Well, typically, as part of our analysis, we would look at the cost/benefit analysis or the risk assessment to study, number one, how they decided—for example, you need a risk assessment, we would assume, to show where you needed to deploy the program. It is at 161 airports, so our question was how did you establish this number? Did you have a risk assessment? And the answer was no. They are in the process of ramping up the program now. Every year, you know, the funding has increased. We assumed that would be justified by a cost/benefit analysis. They don't have one yet, although to their credit they have agreed to complete both a risk assessment and a cost/benefit analysis. But traditionally, we would expect to find that early at program inception, not 4 or five years after you deployed a program.

Ms. EDWARDS. Well, thank you all for your testimony. And Mr. Chairman, I would just say for the record, it would be good to get a cost/benefit analysis and risk assessment before we spend another, you know, $20 million, $2 million, or $2 on the program. Thank you very much.

Chairman BROUN. And I agree with you, Ms. Edwards. Ms. Adams, you are recognized.

Mrs. ADAMS. Thank you, Mr. Chair. The program, Mr. Willis, has been ongoing since 2007? Is that what I heard?

Mr. WILLIS. The validation research study has been ongoing since 2007.

Mrs. ADAMS. A validation research study since 2007. And I heard you say there was no system out there that you could use that was validated or available, is that correct?

Mr. WILLIS. We are unaware of any behavioral-based screening program that is used that has been rigorously validated, yes.

Mrs. ADAMS. What about Israel's program?

Mr. WILLIS. We have not located any study that rigorously tests that.

Mrs. ADAMS. Did they study it?

Mr. WILLIS. We are not provided any information——

Mrs. ADAMS. Did you ask?

Mr. WILLIS. Yes.

Mrs. ADAMS. And they have said they would not provide it?

Mr. WILLIS. We have not been—they didn't say they wouldn't provide it.

Mrs. ADAMS. Okay. So it is maybe the way you were—you asked for it maybe? I am trying to determine, since '07 you have been doing a study. We don't have anything validated. You can't give us a cost/benefit analysis. We are four years out and when you say there is no other programs out there, there are some out there, I believe. Mr. DiDomenica, are there programs out there?

Mr. DIDOMENICA. There are similar programs—excuse me. There are similar programs for behavior assessment, principally for law enforcement. I mean I have been teaching BASS. There is a DHS program called—it is proved by DHS called Patriot. I have another training course called HIDE, Hostile Intent Detection Evaluation. But these programs are given, it may be a few days of training, and then people go off and do their thing. There is no follow up, in other words, how successful it is. I mean people, I think, are getting good ideas, they are getting good techniques, but it is not done in a way where it can be measured and followed up on, and I think that needs to be done.

Mrs. ADAMS. And these programs are all from DHS also?

Mr. DIDOMENICA. There is one that is approved. In other words, it is approved for funding. And—but they are not DHS programs.

Mrs. ADAMS. Okay. So they are funded but they are trying to then—they are kind of sent out and there is no true follow up. Is that what you are saying?

Mr. DIDOMENICA. Yeah, there is no collection of data about success or failures or effectiveness. It is like a lot of law enforcement training, and you are probably aware of this, that you go in for a class, you sit there for a week, you get a certificate, and you walk out the door and that is the end of it. So I think, unfortunately, that just falls in line with a lot of the training that is done. And I think for this program, it is—you know, what is at—for what is at stake, we need to be better at how we follow up on this.

Mrs. ADAMS. I know in my certificate we had to go back for training every so often or else we lost our certificate. So I can relate to having to keep your training and your skills honed. I appreciate that. No more questions, Mr. Chair.

Chairman BROUN. Thank you, Ms. Adams. I want to thank the witnesses for being here today. I appreciate you all's testimony and I appreciate the Members, all the questions that we have had. This is a very interesting topic. I am, again, very disappointed the TSA has refused to come because there are a lot of questions that I know Ms. Edwards and I both would like to have asked TSA if they had graced us with their presence. And hopefully we don't have to go down the road of requiring them to be here in the future. But we will look into that and they will be here at some point, I hope voluntarily. And I hope you will pass that along to the folks that are in the position to make that decision.

Members of the Subcommittee may have additional questions for the witnesses, and we ask that you all will respond to those in writing. The record will remain open for two weeks for additional comments by Members. The witnesses are excused and the hearing is now adjourned.

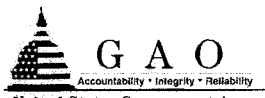
[Whereupon, at 12:00 p.m., the Subcommittee was adjourned.]

Appendix I

Answers to Post-Hearing Questions

ANSWERS TO POST-HEARING QUESTIONS

Responses by Mr. Stephen Lord, Director, Homeland Security and Justice Issues, Government Accountability Office



United States Government Accountability Office
Washington, DC 20548

May 3, 2011

The Honorable Paul Broun, M.D.
Chairman
The Honorable Donna Edwards
Ranking Member
Subcommittee on Investigations and Oversight
Committee on Science, Space, and Technology
House of Representatives

Subject: *Aviation Security: Responses to Posthearing Questions for the Record*

On April 6, 2011, I testified before your committee on the Transportation Security Administration's (TSA) behavior detection program known as Screening of Passengers by Observation Techniques (SPOT). This letter responds to four questions for the record that you posed. The responses are based on work associated with previously issued GAO products.[1] Your questions and my responses follow.

1. **In the GAO report, references are made to TSA relying on unpublished research. Please elaborate on those conversations:**
   a. **How much did TSA rely on unpublished research?**
   b. **Did TSA provide GAO – to its satisfaction – all research documents relied on for implementation of the SPOT program?**

In our May 2010 report[2], we stated that Department of Homeland Security (DHS) Science and Technology officials questioned the findings of a report by the National Research Council, which noted that behavior and appearances monitoring might be able to play a useful role in counterterrorism efforts but stated that a scientific consensus does not exist regarding whether any behavioral surveillance or physiological monitoring techniques are ready for use in the counterterrorist context given the present state of the science.[3] These officials stated that the report did not consider recent findings from unpublished DHS, defense, and intelligence community studies. A DHS Science and Technology program director told us that more recent

[1]See GAO, *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 (Washington, D.C.: May 20, 2010). GAO, *Opportunities to Reduce Potential Duplication in Government Programs, Save Tax Dollars, and Enhance Revenue*, GAO-11-318SP (Washington, D.C.: Mar. 1, 2011).

[2]GAO-10-763.

[3]National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008).

unpublished research sponsored by DHS, the Department of Defense, and the intelligence community is promising in that it has demonstrated some linkages between behavioral and physiological indicators and deception. However, DHS's Science and Technology Directorate could not provide us with specific contacts related to the sources of this research. In its comments on our report, DHS stated that it had provided us with all requested documents that represent DHS's Science and Technology Directorate research. We agree that we received requested DHS Science and Technology documents. However, DHS did not provide us with any contact information for unpublished studies by the Department of Defense and other intelligence community studies that DHS had cited as support for the SPOT program. Without such information, we are unable to verify the contents of these unpublished studies.

In addition, National Research Council officials stated that an agency should be cautious about relying on the results of unpublished research that has not been peer reviewed, such as that DHS stated was generated by DHS and the defense and intelligence community, and using unpublished work as a basis for proceeding with a process, method, or program. Moreover, we have previously reported that peer review is widely accepted as an important quality control mechanism that helps prevent the dissemination of potentially erroneous information.[4]

2. **The GAO report identifies shortcomings with SPOT's data collection and record keeping.**
    a. **While TSA concurred on the need to fix its data problems and expressed willingness to do so going forward, what is the reliability of the data collected thus far?**
    b. **How concerned should members of this Subcommittee be that S&T is relying on that data as the basis for its as-yet-unpublished validation report?**

In its comments on our May 2010 report, TSA stated that the completeness, accuracy, authorization, and validity of data collected during SPOT screening has been greatly enhanced. According to TSA, additional controls have been put in place to address the shortcomings of the previous database. Although we received updates on improvements made to the SPOT database from TSA in late 2010, we have not assessed the reliability of TSA's updated SPOT database.

In our May 2010 report, we identified weaknesses in TSA's SPOT database. We determined that because of these data-related issues, meaningful analyses could not be conducted to determine if there is an association between certain behaviors and the likelihood that a person displaying certain behaviors would be referred to a law enforcement officer or whether any behavior or combination of behaviors could be used to distinguish deceptive from nondeceptive individuals. DHS Science and Technology Directorate recognized weaknesses in the procedures for collecting data on passengers screened by SPOT and planned to more systematically collect data during its study by, for example, requiring behavior detection officers to record more complete and accurate information related to a passenger referral immediately

[4]GAO, *University Research: Most Federal Agencies Need to Better Protect against Financial Conflicts of Interest*, GAO-04-31 (Washington, D.C.: November 2003).

following resolution. However, as we reported in May 2010, if DHS uses operational SPOT data from TSA's SPOT database that was entered prior to March 2010 when improvements to its SPOT database were made, it lacks assurance that SPOT data can be used effectively to help validate the science underlying the program.

3. Should funding for the SPOT program be reduced or eliminated completely until the program is scientifically evaluated?

We have previously reported that the program should not be expanded until the science underlying the program is validated.[5] In our March 2011 report, we reported that validation of TSA's SPOT program is needed to justify continued funding or expansion. Moreover, the results of an independent assessment are needed to determine whether current validation efforts are sufficiently comprehensive to validate the program, and to support requests for increased funding. As such, we suggested that Congress may wish to consider limiting program funding pending receipt of an independent assessment of TSA's SPOT program. We continue to believe that additional increases in program funding should not be provided until DHS has a validated scientific basis for using behavior pattern recognition for observing airline passengers for signs of hostile intent.

4. Does funding for SPOT constitute a practical use of taxpayer dollars?

The nation's constrained fiscal environment makes it imperative that careful choices be made regarding which investments to pursue and which to discontinue. As one layer of aviation security, the SPOT program has an estimated projected cost of about $1.2 billion over the next 5 years. TSA's investment in SPOT could reach about $1 billion by fiscal year 2012. According to TSA, SPOT referrals have led to about 2,100 arrests for offenses such as fraudulent documents, immigration violations, and outstanding warrants. However, questions remain as to whether behavior detection principles can be reliably and effectively used for counterterrorism purposes in airport settings to identify individuals who may pose a risk to the aviation system. SPOT officials also told us that it is not known if the SPOT program has ever resulted in the arrest of anyone who is a terrorist, or who was planning to engage in terrorist-related activity. In addition, as we pointed out in our May 2010 report, TSA deployed SPOT without a comprehensive risk assessment, cost-benefit analysis, or strategic plan that possessed the key characteristics of a successful strategy. Such analyses and plans are important in determining whether a program is viable and cost-effective, and is implemented in a manner that will achieve desired results. As such, we recommended in our May 2010 report that TSA conduct a comprehensive risk assessment to determine the effective deployment of SPOT, perform a cost-benefit analysis, and revise and implement the SPOT strategic plan, among other things. Further, in its comments on our May 2010 report, DHS noted that TSA is developing an initial cost-benefit analysis and that the flexibility of behavior detection officers already suggests that behavior detection is cost-effective. However, it is not clear from DHS's comments whether its cost-benefit analysis will include a comparison of the SPOT program with other security screening programs, such as random screening, or already existing security measures as we recommended. Completing a cost-benefit analysis could provide TSA management with analysis on whether SPOT

[5]GAO-11-318SP.

funding is a prudent investment, as well as whether the level of investment in SPOT is appropriate.

If you have any questions about this letter or need additional information, please contact me at (REDACTED)

Stephen M. Lord
Director, Homeland Security and Justice Issues

Responses by Mr. Larry Willis, Program Manager, Homeland Security Advanced Research Projects Agency, Science and Technology Directorate, Department of Homeland Security

Questions submitted by Chairman Paul C. Broun

Q1. Question: Does S&T's evaluation seek to validate the underlying behavioral indicators that form the basis of the SPOT program?

A1. Response: The scope of the study was to conduct an operational examination of the existing indicators contained within the Screening Passengers by Observational Techniques (SPOT) Referral Report. The results of the study provide evidence to support the criterion-related validity (classification accuracy) of the SPOT Referral Report. In a comparison of Operational SPOT and random screening selection outcomes, the classification accuracy for Operational SPOT was significantly more accurate in identifying high-risk travelers as defined by possession of serious prohibited and illegal items (weapons, fraudulent documents, etc.) and law enforcement arrests. This finding was based upon a comparison of Operational SPOT and random screening at 43 airports for a period of nine months and included over 23,000 Operational SPOT screenings and 70,000 random screenings.

Q2. Question: For the purpose of the S&T study, you describe 'high risk travelers' as "those passengers in possession of serious prohibited and/or illegal items or individuals engaging in conduct leading to an arrest."

a. Why is 'terrorism' not included in the definition of high risk travelers?

A2 a. The number of terrorists identified as traveling through airports is too infrequent to support the inclusion of terrorists as high-risk passengers in an empirical comparative analysis of screening methodologies. In keeping with the best practice of developing proxy measures, the Science and Technology Directorate's study defined high risk travelers using behaviors common to both terrorists and criminals, such as attempting to conceal identity and smuggling of potentially dangerous materials.

b. Has the definition of high risk travelers changed from when SPOT was first implemented? If so, how?

A2 (b.) The definition has not changed.

Q3. At a recent Oversight and Government Reform hearing, TSA stated that it was introducing training for screeners to put travelers at ease while going through screening.

a. What impact would this, and other countermeasures employed by travelers such as training to hide indicators, or anti-anxiety drugs, have on a BDO's ability to identify an individual intending to cause harm?

A2 (a.) Screening of Passengers by Observation Techniques (SPOT) indicators are based on the involuntary physical and physiological behaviors that occur when a person has a fear of discovery. Research supports that these behaviors are difficult to countermeasure. First, involuntary behaviors originate in an area of the brain that individuals do not have control over. People cannot stop these behaviors from occurring; rather they must try to mask or suppress them once they are triggered. Second, nonverbal behavior is more complex and more difficult to control than verbal communication because there are many areas of nonverbal behavior an individual needs to control, such as facial expression, posture, etc. Third, deception is a cognitively demanding state, and this makes body movements even more difficult to control, because people have lower cognitive capacity when they are trying to lie.

Research has not yet examined how medication, surgery, disguise, or drugs affect human behavior in these situations, and this research is needed by the scientific community. Even though medication or drugs may suppress some behaviors and body movements, they may produce other signals to suggest that the person has taken this medication.

Q4. How does TSA ensure that BDOs are using indicators to screen passengers rather than something more troublesome like profiling or racial bias?

A4. Behavior Detection Officers (BDO) and candidates are trained to identify behaviors, and work to resolve any suspicions based on the training protocols. The BDO training distinguishes between subjective profiling and proven scientific methods. They are specifically trained not to consider ethnicity or race-and or other traits that are not associated with behavior. Additionally, BDOs work in teams which aids in integrity. Furthermore, the program office regularly performs Standardization

Visits with refresher training. Finally, the Screening of Passengers by Observation Techniques (SPOT) Transportation Security Managers, who are the first line supervisors to the BDOs, are required to spend time on the floor monitoring the BDOs to ensure they are applying the behaviors in accordance with the SPOT standard operating procedures.

Q5 a. On what basis was the SPOT checklist of indicators selected?

A5 (a.) The behavioral indicators incorporated within Screening of Passengers by Observation Techniques (SPOT) are based on both law enforcement experience and the most recent scientific findings.

Additionally, the work of Dr. David Givens, Director of the Center for Nonverbal Studies, was utilized in selecting the SPOT behaviors. Dr. Givens is recognized as an expert in nonverbal behavior. Behaviors outlined in his Nonverbal Dictionary were selected based on their relationship to stress, fear, and deception cues associated with the fear of discovery and integrated into the SPOT program.

Q5 b. Why doesn't the S&T study evaluate the validity of the indicator list? Do you believe this would be helpful?

A5 (b.) The Science and Technology Directorate's (S&T) study did directly evaluate the indicator list as executed through the existing Screening Passengers by Observational Techniques (SPOT) Standard Operating Procedure (SOP).

Q6. According to the GAO report, S&T officials "agreed that SPOT was deployed before its scientific underpinnings were fully validated." (p. 15). Additionally, in discussing the S&T study, the GAO report states, "S&T's current research plan is not designed to fully validate whether behavior detection and appearances can be effectively used to reliably identify individuals in an airport terminal environment who pose a risk to the aviation system." (p. 20). Additionally, in the first paragraph of Dr. Maria Hartwig's written testimony, she says, "In brief, the accumulated body of scientific work on behavioral cues to deception does not provide support for the premise of the SPOT program. The empirical support for the underpinnings of the program is weak at best, and the program suffers from theoretical flaws."

a. Prior to implementing SPOT, why did TSA not validated the science behind the program?

A6 (a.) Prior to the Transportation Security Administration's Screening of Passengers by Observation Techniques (SPOT) program, no behavior-based program had ever been rigorously scientifically validated. The program was established on widely accepted principles supported by leading experts in the field of behavioral science and law enforcement.

b. Why did the S&T validation study not validate "whether behavior detection and appearances can be effectively used to reliably identify individuals in an airport terminal environment who pose a risk to the aviation system?"

A6 (b.) The Science and Technology Directorate (S&T) sponsored study did directly examine the extent to which "behavior detection and appearances," as represented in the existing Screening Passengers by Observational Techniques (SPOT) indicators, can be effectively used to identify high-risk travelers, which is an examination of classification accuracy (criterion-related validity). Results of the study found support for criterion-related validity; that is, there is evidence that the SPOT indicators are accurate in identifying outcomes and is significantly more accurate in doing so than random screening.

c. How do you respond to Dr. Hartwig's comment?

A6 c. During the recent testimony, Dr. Rubin responded to a similar question by stating that the published research literature on the link between behavioral, physiological, and verbal cues to deception and general suspicious behaviors is mixed, rather than non-supportive as represented by Dr. Hartwig. The Science and Technology Directorate (S&T) agrees with Dr. Rubin's assessment.

Q7. Who originated the SPOT program, was it Carl Maccario, as Dr. Ekman states in his written testimony, or was it Lieutenant DiDomenica, who says his PASS program was the basis for SPOT? Response: After the terrorist attacks of 9/11, behavior recognition and analysis concepts were adapted and modified by the Massachusetts State Police (MSP) Troop F (Lieutenant DiDomenica) assigned to Boston Logan International Airport (BOS). Their program was modified to meet the legal, social, political, financial, and resource limitations of the United States and was merged with drug interdiction techniques used by United States

law enforcement. MSP named this program Behavior Assessment Screening System and trained all law enforcement officers assigned to BOS in its use as an enhanced security measure to the newly instituted security checkpoint screening system of the Transportation Security Administration (TSA).

The Screening of Passengers by Observation Techniques (SPOT) program was developed by TSA (Carl Maccario), with assistance from MSP, to meet TSA-specific security and public service needs, with particular emphasis on the protection of individual civil rights, privacy, and to mitigate possible complaints of racial profiling.

a. What role did the Israeli model play?

A7 (a.) The SPOT subject matter expert was initially trained in Israeli Behavior Pattern Recognition (BPR). Many of the BPR concepts are contained in SPOT such as informally interacting with passengers who are in line at the security checkpoint queue.

b. What aspects of the Israeli model are based on behavioral science?

A7 (b.) TSA defers to the Government of Israel to respond as appropriate, as they are the subject matter experts on their security model.

Q8. Dr. Ekman distinguishes his experiments from those of his critics by emphasizing that his focus is on "high stake lies, in which the person lying has a lot to gain or lose by success or failure." He specifically addresses the work conducted by Dr. Hartwig, stating, "She has dealt with low-not-high-stake lies which have little relevance to my work or to the situation faced in SPOT." Conversely, Dr. Hartwig states, "Neither the research in general nor specific results on high-stake lies support the assumption that liars leak cues to stress and emotion, which can be used for the purposes of lie detection."

a. Given these opposing views, what is your assessment?

A8. As Dr. Rubin stated during his testimony, the published research literature is mixed on the topic of behavioral, physiological, and verbal cues to deception and general suspicious behaviors. Ideally, one might expect greater consensus and support from the academic research base prior to fielding a screening program; however, academic research alone is insufficient. Once a screening program is fielded, regardless of how supportive the academic research base may be, prudent research requires the conduct of operational experiments to validate the effectiveness of the screening program and if effective, to then conduct additional research to optimize its effectiveness. The reality is that behavior-based screening is currently used operationally by DHS, the U.S. Department of Defense, the U.S. intelligence community, law enforcement, and by numerous other countries. Increased focus should be applied to conducting field research on these programs.

Q9. Please indicate each and every research effort that the DHS Science & Technology Directorate (S&T) is conducting on behalf of the Transportation Security Administration (TSA). This should include all efforts the S&T Directorate is taking on behalf of TSA and not simply be limited to work that S&T is performing regarding the TSA SPOT program.

Please include in this list the following information:

- *The name of the TSA effort DHS S&T is supporting.*
- *The purpose of the S&T research or task.*
- *The amount of financial reimbursement S&T is receiving from TSA for each effort.*

A9. The Science and Technology Directorate (S&T) partners with the Transportation Security Administration (TSA) on several research and development tasks. Below are the projects and associated funding from FY 2010 reimbursed by TSA: (*NOTE: * indicates projects are funded by TSA and do not appear in S&T budget documents)*

Project Name: Secure Carton
Financial Reimbursement from TSA: N/A
Description: Develop (at the request of TSA and DHS Policy) a shipping carton embedded with security sensors that detects tampering or opening of the carton once closed. It is scalable and applicable across various shipping modalities, including maritime and air cargo, and can communicate a tamper event of the internal cargo to a radio frequency identification reader, when interrogated. The interaction with TSA has been to keep them informed of the project. S&T intends to test the product for inclusion on the TSA qualified products list. Secure Carton is a Phase-III Small Business Innovation Research (SBIR) - Phases I &

II were funded by S&T SBIR Program and Phase III was funded with S&T Borders and Maritime Security Division FY09/10 project funds.

Project Name: Secure Wrap
Financial Reimbursement from TSA: N/A
Description: Secure Wrap is being developed for TSA and DHS Policy. It is a flexible wrapping material that provides a visible indication of tamper evidence and can be deployed with little to no change to current supply chain logistics and processes. The interaction with TSA has been to keep them informed of the project. S&T intends to test the product for inclusion on the TSA qualified products list. Secure Wrap is a Phase-II SBIR with all funding provided by DHS S&T SBIR Program.

Project Name: Autonomous Rapid Facility Chemical Agent Monitor Project
Financial Reimbursement from TSA: N/A
Description: Develop a low-cost, fully autonomous, chemical vapor monitor that is intended to "detect-to-warn" of the presence of up to 17 chemical warfare agents and high-priority toxic industrial chemicals within a single device at both immediately dangerous to life and health and permissible exposure limit concentrations. The monitor will be able to operate continuously in closed or partially enclosed facility 24hrs/day, 365 days/yr.

Project Name: Chemical Security Analysis Center (CSAC) Project
Financial Reimbursement from TSA: N/A
Description: Develop and sustains expert reach-back capabilities to provide rapid support in domestic emergencies. The CSAC serves as the Nation's first centralized repository of chemical threat information (hazard and characterization data) for analysis of the Nation's vulnerabilities to chemical agent attacks. To ensure a cohesive effort to evaluate threats and countermeasures, CSAC conducts key analytical assessments, such as material threat assessments (MTAs), hazard assessments, and the Chemical Terrorism Risk Assessment (CTRA). The DHS Office of Infrastructure Protection, Office of Health Affairs, TSA, and Intelligence & Analysis are the primary DHS customers for CSAC products. CSAC provides completed MTAs to Health and Human Services to fulfill BioShield requirements.

Project Name: Model Large-Scale Toxic Chem Transport Release Project
Financial Reimbursement from TSA: $800,000
Description: Focus on developing an improved understanding of large-scale releases of toxic inhalation hazards. Aspects of the project include improved modeling, first responder procedures, and industrial safety in addition to the development of enhanced mitigation strategies.

Project Name: Canine Detection R&D Project (FY10)
Financial Reimbursement from TSA: N/A
Description: Assess the performance of TSA certified explosive detection canine teams when screening air cargo. This effort is in support of the TSA National Explosives Detection Canine Team Program (NEDCTP) effort to independently test performance measures in operational environments in order to make decisions on concepts of operations. Independent experts collect and present the data from canine operational assessments and make recommendations on canine training or deployment to optimize canine explosives detection.

Project Name: Homemade Explosives (HMEs) Stand Alone Detection Project (FY10)
Financial Reimbursement from TSA: N/A
Description: Identify, evaluate, and improve HME detection technologies and screening methods through the collection and analysis of detection data and images from a wide variety of commercial off-the-shelf (COTS) explosive detection systems (EDS), computed tomography, and x-ray diffraction equipment. This helps TSA determine how to improve screening system performance through hardware and software (image processing) upgrades. In addition, this project evaluates COTS explosives detection equipment in laboratory settings to determine detection limits, false-alarm rates, and documents unique homemade explosive (HME) properties for detection exploitation.

Project Name: Air Cargo Project (FY10/FY11)
Financial Reimbursement from TSA: FY 10 $1.1 million
Description: Identify and develop next generation screening systems to mitigate the threat of explosives placed in air cargo containers. Activities include developing technologies to enable more effective and efficient air cargo screening (in-

cluding break-bulk, palletized, and containerized configurations screening) with reduced operational costs and false-alarm rates.

Project Name: Algorithm and Analysis of Raw Images (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Develop a non-proprietary database of explosive-detection images which will be provided to all detection-program participants. Collect and consolidate images, including those of novel explosives, from commercial vendors and coordinates the purchase of additional images and data from computed tomography, explosive detection systems, trace, emerging devices and other technologies. The evaluation of these images will help determine the causes of false alarms over many types of scanning systems.

Project Name: Automated Carry-On Detection (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Develop advanced capabilities to detect explosives and concealed weapons in carry-on luggage. This project also will introduce new standalone or adjunct imaging technologies, such as computed tomography, to continue the improvement of checkpoint detection performance and the detection of novel explosives.

Project Name: Automated Threat Recognition (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Develop and evaluate automated target recognition algorithms for advanced imaging technology in a test bed with the goal to automatically and reliably detect threats on passengers, eliminating the need for human interpretation in order to improve detection and false alarm performance and reduce privacy concerns. The December 25, 2009 incident clearly shows the importance of detecting threats hidden on passengers' bodies. This research will guide further enhancements necessary to reach full-scale development and deployment.

Project Name: Detection Technology and Material Science (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Evaluate advanced detection algorithms, improves explosives detection and develops and tests advanced materials for trace sample collection.

Project Name: Explosives Trace Detection (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Develop advanced capabilities to detect explosives (including homemade explosives) through improved trace sampling and detection technologies. Develops trace detection standard materials that can be used as field performance standards for deployed trace detection systems. Characterizes trace explosives chemical and physical signature properties to inform advanced trace detector system design.

Project Name: Checked Baggage (FY10/FY11)
Financial Reimbursement from TSA: FY 10 $5.5 million
Description: Drive commercial development of next-generation systems that will substantially improve performance and affordability of checked baggage screening. Commercial development is driven when the test results referred to below are incorporated into TSA's increased performance requirements for screening systems. Vendors must then meet these requirements for consideration during TSA acquisition. Test and evaluation of these systems will focus on probability of detection, number of false alarms, and throughput. The project also measures affordability of these systems by evaluating initial purchasing cost, operating costs, maintainability, and other elements of the full life-cycle costs.

Project Name: Mass Transit (formerly Suicide Bomber) (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Identify the infrastructure characteristics and security concept of operations for surface transportation systems in order to drive a security technology development strategy designed to combat the explosive threat within the operational requirements of the transportation systems. Assessments will be conducted at transit authorities to frame the technology development solution space. Currently fielded technologies will be evaluated for potential enhancement.

Project Name: Next Generation Passenger Checkpoint (FY10/FY11)
Financial Reimbursement from TSA: FY 10 $2.1 million
Description: Develop the next-generation detection system architecture to screen passengers for explosives at aviation checkpoints. This project also investigates new emerging liquid- and gel-based explosive threats and includes them in a comprehensive detection system.

Project Name: Predictive Screening Project
Financial Reimbursement from TSA: N/A
Description: Derive the observable behavioral indicators and develops technologies to automatically identify, alert authorities to, and track suspicious behaviors that precede suicide bombing attacks. The Science and Technology Directorate will test technologies at ports-of-entry, transit portals, and special events.

Project Name: Aircraft Vulnerability Tests (FY10/FY11)
Financial Reimbursement from TSA: FY10 $6.6 million
Description: Assess the vulnerability of narrow- and wide-body aircraft passenger cabins and cargo holds to explosives. These vulnerability assessments will analyze blast/damage effects of explosives and determine the minimum threat mass required to cause catastrophic damage to various aircraft types. The assessments will also identify the detection limits for bulk screening systems. Develop and assess hardened unit load devices (HULDs) for blast mitigation in air cargo. These HULD development efforts will provide reduced weight air cargo containers for blast protection while minimizing impact on commerce.

Project Name: Homemade Explosives (HME) Characterization (FY10/FY11)
Financial Reimbursement from TSA: N/A
Description: Determine the impact, friction, and electrostatic-discharge sensitivities of HME threats. This data facilitates the safe handling and storage of HME materials during research and development activities. Technology efforts to identify, evaluate, and improve HME detection technologies and screening methods through the collection of raw data and images from a wide variety of commercial off-the-shelf (COTS) explosive detection systems (EDS), computed tomography, and x-ray diffraction equipment are also conducted. This helps TSA determine how to improve EDS performance through hardware and software (image processing) upgrades. In addition, this project evaluates COTS equipment in laboratories to determine detection limits, false-alarm rates, and documents unique HME properties for detection exploitation.

Project Name: Facility Restoration Demonstration Project
Financial Reimbursement from TSA: N/A
Description: Develop a systems approach to response and recovery of critical transportation facilities following a chemical agent release. This project develops remediation guidance, efficient pre-planning tools, identifies decontamination methods, identifies sampling methods, and develops decision analysis tools.

Project Name: Operational Tools for Response and Restoration Project
Financial Reimbursement from TSA: N/A
Description: Develop a suite of state-of-the-science indoor-outdoor predictive tools to characterize the extent and degree of biological contamination, incorporating the best-available deposition, degradation, and surface viability data. This project will provide validated interagency sampling plans and improved statistical sampling design to support characterization and decontamination planning.

Project Name: Bridge Vulnerability Project
Financial Reimbursement from TSA: None
Description: Develop an understanding of the vulnerabilities of different types of bridges to terrorist threats. This project will evaluate vintage bridge components to improve understanding of explosives effects and to refine blast modeling tools. The approach is unique in that it examines actual bridge sections exposed to wear or aging instead of fabricated specimens. As a result, it will provide more accurate vulnerability information for aging bridges and allow for refinement of existing numerical models that predict failure of bridge components. The project is using the Golden Gate Bridge, Crown Point Bridge (New York State - Lake Champlain), and Manhattan Bridge (New York City East River), and the Fort Steuben Bridge (Ohio) for homeland security research on potential effects of an improvised explosive device (IED) attack and other plausible threats against a bridge. These efforts are in partnership with the Maine Department of Transportation (DOT), NY DOT, NYC DOT, Ohio DOT, Golden Gate Bridge Authority, and the Federal Highway Administration.

Project Name: Blast/Projectile – Protective Measures and Design Tools
Financial Reimbursement from TSA: None
Description: Identify and evaluate protective measures and design guidance for protecting the Nation's most critical infrastructure assets. The project considers novel materials, design procedures, and innovative construction methods to aid in constructing or retrofitting infrastructure. This will numerically analyze pro-

tective designs against blast and projectile threats and conduct physical demonstrations to assess effectiveness.

Project Name: Advanced Incident Management Enterprise System (AIMES)
Financial Reimbursement from TSA: None
Description: Develop the next-generation incident-management enterprise system and builds upon the Unified Incident Command and Decision Support architecture and Training, Exercise & Lessons Learned framework. This will integrate all elements of the incident management enterprise to provide a secure, scalable, interoperable, and unified situational awareness to the responder community.

Project Name: Rapid Mitigation and Recovery Project
Financial Reimbursement from TSA: None
Description: Investigate, assess, and develop candidate technologies and methodologies that will reduce or eliminate the release of toxic inhalation hazard (TIH) from the two threat scenarios of interest (.50 caliber AP and small IED). Assess potential TIH mitigation technologies, to include development of interface documentation to ensure that identified technologies can be integrated into any existing and or future rail car design efforts. Mitigation technologies and approaches to be assessed include: Self-sealing Technologies and Blast and Fragment Penetration Resistant Materials.

Project Name: Blast Projectile-Advanced Materials Design
Financial Reimbursement from TSA: None
Description: Assess the risk to a tunnel or mass transit station due to a terrorist attack that has the potential of causing catastrophic losses (fatalities, injuries, damage, and business interruption). Information from Integrated Rapid Visual Screening Tool (IRVS) can be used to support higher level assessments and mitigation options by experts. In coordination with TSA, IRVS for Mass Transit Stations and Tunnels were tested in various cities: Boston (Boston Massachusetts Bay Transportation Authority (MBTA), Cleveland, St. Louis, and others. TSA will use the tool to enhance risk assessments of transportation hubs around the country. In addition to TSA, potential users include Office of Infrastructure Protection, Federal Emergency Management Agency, Commercial and Government Facilities, State and local governments, code officials, associations of engineers and architects, the design and construction industry.

Project Name: Community Based CIP Institute
Financial Reimbursement from TSA: FY11 $1million
Description: The shipment of hazardous materials provides a significant target for terrorists. The ability to track hazardous materials (HAZMAT) shipments on a real-time basis is essential for providing an early warning of an impending terrorist threat. The University of Kentucky (UK) will design and organize a functional prototype of a HAZMAT truck tracking center. This project supports a Transportation Security Administration (TSA) program that tracks motor carrier shipments of security-sensitive materials. Collaborating with UK on the project are Morehead State University, Coldstream Digital and General Dynamics Advanced Information Systems. The prototype software is integrated with "smart truck" technology and will contain operational components that will integrate reporting and shipping information with a real-time tracking and situation display capability.

Project Name: Suspicious Activity Reporting Project
Financial Reimbursement from TSA: None
Description: S&T is developing an enhanced analytical tool prototype for the Federal Air Marshal Service (FAMS), Investigations Division. This application, now named iConnex, is a suite of analytical tools that allows investigators to search, find, explore, link, visualize and understand relationships within Suspicious Activity Reports and other law enforcement data sets. The iConnex application is under development using predominantly open-source technologies. The application's architecture targets the technical needs of the law enforcement community by being able to work with an array of structured and unstructured data. The system is designed to be user friendly, and does not require extensive training or support to reach operational capabilities. Once completed, iConnex will be made available to any DHS component or law enforcement agency as a cost-free Government Open Source solution.

Project Name: Law Enforcement Data Fusion
Financial Reimbursement from TSA: None

Description: The Science and Technology is working with Federal Air Marshal Service (FAMS), Investigations Division to develop a geospatial predictive analytics product that will detect, forecast, and disrupt future terrorist attacks and criminal activity - leveraging predictive analytic algorithms and software developed for the Department of Defense community that successfully 'forecast' improvised explosive device locations in Iraq and Afghanistan. This capability will provide FAMS with actionable guidance on the most effective location and allocation of agents to place on high risk flights as well as providing them with increased knowledge of the tactics and procedures of the adversary. This effort utilizes a cloud-computing environment in which national data (Homeland Security Infrastructure Protection Gold, among others) are being brought together and analyzed to support the FAMS mission to discern threats and forecast the location of attacks. As this technology matures at FAMS, the final product will be made available to any DHS component or law enforcement agency as a cost-free Government Open Source solution.

Project Name: Cross-Cultural Validation of Screening of Passengers by Observation Techniques (SPOT)
Financial Reimbursement from TSA: N/A
Description: Provide empirical validation of existing behavioral indicators employed by DHS' operational components to screen passengers at air, land, and maritime ports, including those indicators contained within TSA's SPOT. This effort complements the automated prototype work and supports development of an enhanced capability to detect behavioral indicators of hostile intent at a distance. The project will integrate these validated behavioral indicators into the screening concept of operations through each component's existing training programs.

Project Name: Future Attribute Screening Technologies Mobile Module (FAST M2)
Financial Reimbursement from TSA: N/A
Description: Develop a prototype screening facility containing a suite of real-time, non-invasive sensor technologies to detect behavior indicative of malintent (the intent or desire to cause harm) rapidly, reliably, and remotely. The system will measure both physiological and behavioral signals to make probabilistic assessments of malintent based on sensor outputs and advanced fusion algorithms. Federal, state, and local authorities may use the fully developed FAST system in primary screening environments to increase the accuracy and validity of people screening at special events, airports, and other secure areas. FAST will measure indicators using culturally independent and non-invasive sensors. FAST will use an ongoing, independent peer review process to ensure objectivity and thoroughness in addressing all aspects of the program.

Project Name: Hostile Intent Detection - Automated Prototype
Financial Reimbursement from TSA: N/A
Description: Develop real-time, non-invasive, and culturally independent, hostile-intent detection video extraction algorithms to identify unknown or potential terrorists through an interactive process.

Project Name: Human Systems Research
Financial Reimbursement from TSA: FY10 $1.7 million
Description: Examine ways to maximize human performance across DHS end-user tasks and activities. Activities under this project include research on exceptionally performing (EP) screeners, development of a human factors research roadmap, a study of airport dynamics and the development of a cognitive assessment tool.

***Project Name:** Aviation Security Enhancement Partnership (ASEP) Evaluating TSA's Comprehensive Airport Security Strategy
Financial Reimbursement from TSA: FY10 $1 million
Description: The project will deliver an evidence-based assessment and a research design for a comprehensive evaluation of the efficacy of the Transportation Security Administration's Playbook to ensure that it has the intended prevention and deterrent effects in and around U.S. airports.

***Project Name:** Intelligent Closed Circuit Television (iCCTV) Project
Financial Reimbursement from TSA: FY10 $400,000
Description: Design and construct a data video collection, storage, and distribution capability to support off-line behavioral analysis. The resulting analysis will support an inter- and intra-reliability assessment of the SPOT indicators.

***Project Name:** Behavior Detection Officer (BDO) Selection Instrument Validation Project
Financial Reimbursement from TSA: FY09 $1.25 million (still being completed)
Description: Design and validate a personnel selection instrument to support the hiring of TSA BDO.

Responses by Dr. Paul Ekman, Professor Emeritus of Psychology,
University of California, San Francisco,
and President and Founder, Paul Ekman Group, LLC

Questions submitted by Chairman Paul Broun

Q1. A Nature article from May, 2010 states that you no longer publish all of the details of your works in peer-reviewed literature because those papers are closely followed by scientists in countries such as Syria, Iran and China, which the United States views as a potential threat. A great deal of security related research is conducted in the country in a manner that follows both the principles of peer review as well as the security classification systems Is your work unique in this regard?

A1. I have not done classified research, and I don't know how those who do such research handle the matter of publishing their findings, or any part of their findings. I have been told that classified research is not published, but that is hearsay. Regarding our own research findings, 95% of what we call hot spots -- behaviors which indicate that full disclosure has not occurred -- has already been published in scientific journals or book chapters. We have chosen not to publish a few new findings on hot spots in an attempt not to disclose to potential and actual enemies of our country everything we have found. If we choose to publish a study and it contains these undisclosed hot spots, then we exclude those undisclosed hot spots from the statistical analyses that we do report. Since the incidence of these undisclosed hot spots is quite low, it has not changed the overall findings. Thus we are able to publish on the incidence of 95% of hot spots, and keep to ourselves and those we teach in law enforcement and national security, knowledge of the new unpublished hot spots.

Q2. On pages five and six of your written testimony, you reference a couple of unpublished studies spearheaded by Dr. Mark Frank, one of which you claim shows "behavioral markers can be useful even in situations where the person has yet to commit an illegal act." Did you share any preliminary results from these studies with either TSA or S&T?

A2. The TSA was fully informed of Dr. Frank's study that showed it was possible to detect from hot spots whether or not a person had decided to lie. Past research had focused on identifying lies about behavior that already had occurred. This study showed it was also possible to detect lies about the future intent to engage in a malfeasant action.

Q3. On page seven of Dr. Hartwig's testimony, she responds to your claim from a New York Times interview of being able to teach lie detection "to anyone with an accuracy rate of more than 95 percent." She goes on to say, "However, no such finding has ever been reported in the peer-reviewed literature. More broadly, there is no support for the assertion that training programs focusing on identifying facial displays of emotions can improve lie detection accuracy. How do you respond to those observations?

A3. Dr. Hartwig has made a mistake in what she claims I said, one of many mistakes in her testimony. What I said was that through time-consuming, careful behavioral measurement we have been able to reach accurate determination of who is lying with up to 95% accuracy, but this included combining some physiological measures as well. I also said that we teach law enforcement and national security personnel about our findings, attempting to train them to be able to use our findings in their evaluations without doing the actual time-consuming research. We have not claimed that those we train reach a 95% accuracy level of correct judgments in their work place after our training. We receive reports that they have benefited, and we have a paper under review by a scientific journal that shows that teaching individuals to recognize micro expressions improves their ability to judge the true emotional state of people who are lying. This in combination with a number of published studies (once again not cited or not known by Dr. Hartwig) -- Ekman & O'Sullivan, 1991; Frank & Ekman, 1997; Warren, Schertler & Bull, 2008 - which show a correlation between accuracy at detecting micro expressions and accuracy at detecting lies. But this is found only when the lie is about something the person cares about and there is a threat of considerable punishment if detected.

A meta analysis by Frank & Feeley (2003) and later updated by O'Sullivan, Frank & Hurley (2011) on all the published research examining whether training improves the ability to detect lies, found significant improvements as a result of training. Dr.

Hartwig did not know or chose not to mention these studies which directly contradict her testimony.

The only study which evaluated training in actual real world high stakes security contexts is the new American Institute of Research (AIR) report. The training the SPOT personnel received whose decisions were found to be highly accurate in the AIR study included our training materials, and some of the SPOT personnel were trained by us. Our training is not limited to the face, but includes all of demeanor - gesture, gaze, voice, and speech as well as facial actions.

Q4. You claim SPOT needs more funding and BDOs need more training.

a. How much funding is enough for SPOT?

b. How much training time would you devote to BDOs?

A4 a. I believe SPOT needs to have its personnel observing line of traffic at all major airports. I believe our country would be safer if there were also SPOT personnel at all feeder airports, as the 9/11 hijackers boarded and went through security at feeder airports. The information I have received is that there are no SPOT personnel at feeder airports, and only enough personnel to conduct surveillance at half the lines of traffic at our major airports. I believe this is a terrible mistake, especially given the fact that recruiting and training enough SPOT personnel to have this layer of security in place at all airports would cost less than 1% of last year's DHS budget.

Although I am not fully informed of the changes in the program now underway I believe they include increased training time and more selective recruitment.

A4 a. Regarding training time, since the costs of training are low and the costs of just one terrorist being missed are very high, I believe it merits overkill. I expect that 40 hours of training, spread over a few weeks, would be of benefit. But that is a guess as there is no research available to determine when adding training time stops producing benefits.

There are many questions that could be answered by doing research to find out how many BDOs are needed to cover a given area, what breaks are needed and when to optimize performance, and are people missed who show many of the behaviors on the SPOT checklist.

Q5. What steps should TSA have taken prior to implementing the SPOT program nationwide?

A5. I believe TSA took the appropriate steps: it found out what the Israelis were doing; and it obtained the help and advice from those scientists who had done research relevant to its objectives, not just my work. By the time TSA consulted with Israel about their training, we had already provided training to the Israelis. It should be clear that the training included but was not limited to micro expressions. In our research we measure and find useful hot spots shown in gesture, voice and speech itself. And these too are included in TSA's behavioral profiling.

I believe TSA made the right judgment in adding this layer of security prior to research about how effective it would turn out to be in catching malfeasants. The recent AIR study showed it is effective, but it would have been a mistake, in my judgment, not to have provided the American people with this layer of security before that study was performed.

I regret that the American people are not now being provided with all the layers of security which are available in England and Israel, because there simply are not enough trained Behavior Detection Officers.

**Professor Mark Frank, SUNY Buffalo contributed to some of these responses.*

References

Ekman, P. & O'Sullivan, M. (1991) Who can catch a liar? American Psychologist, 46(9), 913-920.

Frank, M.G., & Ekman, P. (1997) The ability to detect deceit generalizes across different types of high-stake lies. Journal of Personality and Social Psychology 72, 1429-1439.

Frank, M.G, Feeley, T.H., Paolantonio, N. & Servoss, T. J. (2004). Individual and Small Group Accuracy in Judging Truthful and Deceptive Communication. Group Decision and Negotiation 13(1), 45–59.

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. Law and Human Behavior 33(6), 542–543.

Warren, G., Schertler, E., Bull, P. (2008) Detecting Deception from Emotional and Unemotional Cues. Journal of Nonverbal Behavior 33, 59–69.

Responses by Dr. Maria Hartwig, Associate Professor, Department of Psychology, John Jay College of Criminal Justice

Questions submitted by Chairman Paul Broun

Q1. Are there any differences in the behavioral cues associated with a liar being deceitful and the behavioral cues associated with a truth-teller stressed about being perceived as a liar? In other words, how would one distinguish a liar from a truthful person who's afraid of not being believed?

A1. In a situation where liars fear detection, and truth tellers fear not being believed, the behavioral patterns of the two are likely to be very similar. Research supports this, by showing that when liars and truth tellers are highly motivated to be believed, they both display patterns of behavior that are likely to attract deception judgments. That is, they may both show signs of stress and fear; signs which an observer may interpret as indicative of deception. Simply put, it is very difficult, if not impossible, to distinguish between the behavioral signs of stress of a liar who fears exposure and those displayed by a truth teller who fears misjudgment.

Q2. Your testimony talks about a paradigm shift in the approach to lie detection that involves, "moving from passive observation of behavior to the active elicitation of cues to deception." Unlike the Israeli process, BDOs in the U.S. can't realistically stop and interview each passenger several times prior to boarding - how do you propose TSA incorporate this mentality into SPOT? Should it? Is it practical?

A2. It is true that it may not be feasible to interview every single passenger due to the high volume of travelers in the U.S. My suggestion is that the TSA, with the help of an independent panel of experts, should review theories and empirical findings on the elicitation of cues to deception, and entertain the possibility of incorporating some of these methods in their protocol for verbal interactions with travelers. Some form of screening is most likely necessary in order to select passengers for additional scrutiny in the form of questioning. Whether the SPOT method should be used for this screening ultimately depends on the findings of the validation study, which, to my knowledge, has yet to be released.

Q3. What steps should TSA have taken prior to implementing the SPOT program nationwide?

A3. It would have been beneficial to create and consult with a panel of independent experts in the relevant areas, in order to ensure that the procedures are in line with the scientific evidence. Moreover, it is my view that the TSA should have carried out a validation study prior to implementing the program nationwide. Again, a panel of experts could have been of assistance in designing and executing such a validation study.

Responses by Dr. Philip Rubin, Chief Executive Officer, Haskins Laboratories

Questions submitted by Chairman Paul Broun

Q1. What are the challenges that scientists need to address in order to conduct research in an operational setting? 1b. Can these hurdles be overcome?

A1. There are numerous challenges related to conducting research in operational settings. I would like to focus on two of these.

1. Evaluation and analysis both in the laboratory and in the field must be based on specific, testable hypotheses that derive from premises that are established in some sort of orderly and/or rational manner. For example, using voice stress analysis (VSA) to illustrate this, it is essential to first understand what is being measured (that is, what is the specific definition of "voice stress") and understand how these measures might related to outcome measures. In addition, in order to isolate critical variables so that then can ultimately be validated (in the lab or in the field), we also need to consider potential interactions of variables that might affect results and other factors that could bias or shape experimental results, including any critical contextual considerations. In the case of approaches like VSA, field tests should not be conducted prior to demonstrating a valid and reliable approach for characterizing and quantifying, if possible, the underlying variables. Once these have been established, it is then possible to move to the field. If the premises are weak or cannot be established, there is little point in moving to field evaluation.
2. Laboratory studies have the advantage that they often provide for the ability to precisely control experimental conditions. The disadvantage is that they often lack what is sometimes called "ecological validity." That is, what is being measured in the laboratory may not accurately capture the phenomena that you are trying to study, often because critical contexts have been removed. Field evaluation lets you study events in their natural environment. This has been standard in the ethological approach and in many other instances including primate research, research on children, and research in organizational and institutional settings. Unfortunately, with this greater realism sometimes comes a consequent loss of experimental control.

Overall, the best approach would be to first clearly nail down a good, concrete understanding of critical variables and the premises that give rise to them. These should be experimentally evaluated and understood prior to field evaluation. An assessment of potentially critical contextual variables is also essential. At that point (but not until then), field evaluation is possible and can provide a rich and realistic approach for evaluating data and programs. Although there are often limitations in the field, clever and informed experimental design can go a long way to assisting with the design of studies that have great utility. If they cannot be used to fully study a system, they can often be informative and useful as they relate to aspects of the problem.

Q2. (Regarding the comments of Dr. Ekman and Dr. Hartwig). Given these opposing observations, what is your analysis?

There appears to be very little in the peer-reviewed, scientific literature to help differentiate high versus low-risk lying and their relationship. As both Dr. Ekman and Dr. Hartwig have indicated, research is needed in this area. Peer-reviewed research would be the useful to establish and solidify scientific validity of results. Such work can be done without jeopardizing security.

Q3. . . . what thoughts do have on the manner in which the SPOT program was implemented?

A3. As you have noted, I agree with Dr. David Mandel's comments from the summary of the NRC workshop that I chaired, called "Field Evaluation in the Intelligence and Counterintelligence Context: Workshop Summary".

> "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."

Unfortunately, it appears that the SPOT program was implemented before its underlying premises, measures, indicators, etc., could be adequately scientifically evaluated and, if necessary, validated in even a remotely meaningful way. Instead, they appear to have been rushed into the field due to a combination of fear, zeal, passion, folklore, intuition, and enthusiasm about controversial scientific results, such as "micro-expressions." As of the time of the April 6, 2011 hearing, and the end of my contribution to the TAC report, I had not been provided with information about the "indicators" used in the SPOT program, so I can only speculate about them. However, if they were things like facial micro-expressions, behavioral indicators such as gaze direction or head tapping, etc., then they should all be subject to scientific scrutiny. Why are such measures being selected? What is the current state of scientific knowledge regarding their validity? If little is known about them, can then be evaluated scientifically? If not, then they should not be used. On other possible measures such as excessive sweating, aberrant behavior, etc., it would be useful to understand the science on how these behaviors related to outcome measures. For example in voice stress analysis (which does not appear to be a reliable measure) which is supposedly related to changes in voice "micro-tremors", is the appropriate indicator greater or smaller magnitude of micro-tremor?

Given the enormous stakes related to national security in transportation, and also to work done by our intelligence and counter-intelligence communities, my strongest recommendation for the Committee would be that the money currently being devoted to (and in my opinion wasted on) this program should immediately be redirected to a *large-scale effort to solicit the best possible scientific and technical guidance related to the detection of deception using behavioral indicators.* The end product should include a clear statement of what works, what does not, what remains controversial, and how to move ahead. The TAC did not have the independence, expertise, breadth of knowledge, nor latitude to take on this challenge, not was it asked to do so. Such a study should be broader than SPOT and should include considerations of approaches like voice stress analysis, facial expression, remote physiological monitoring, and neuroimaging. Members of such a group should have expertise in physiology, behavioral science, psychology, neuroscience, linguistics, statistics and methodological design, and related areas. It is essential that any group working on such a project be independent of DHS and TSA. Scientific evaluation of programs like SPOT and other programs related to the detection of deception can be done in a manner that does not provide unique knowledge to those who would wish to harm us.

Q4. How do you respond to DHS' preliminary assertion that SPOT is significantly more effective than random screening?

A4. As a member of the Technical Advisory Committee I would have to say that this assertion on the part of DHS is not a meaningful or useful one. The base rate for outcomes is too small to be statistically reliable and/or meaningful. If DHS is making an assertion of this sort, then they need to more clearly define and quantify what "significantly more effective than random screening" means. In a population of 100,000 events are 2 observations significantly different than 1? How about 3 versus 1? Or 100 versus 1? What does significance mean as DHS is using the term and what do they mean by "effective"? Small numbers in large populations can be meaningless and simply part of the randomness and background noise that normally occur in most systems. Given the controversial and costly nature of this program, scientific and statistical rigor should be essential. I find such a statement to be misleading and potentially dangerous. Politicians, policymakers and the lay public, will hear something like "SPOT is significantly more effective than random screening" and may assume that this program is effective, useful, and has been adequately scientifically evaluated. To this point the effectiveness and usefulness have not been established. The scientific evaluation has been inadequate and has not been approached in a manner that would lead to greater knowledge regarding the program. Establishing scientific credibility has the potential to be helpful to programs of this sort, but that requires full, well thought out, independent, credible, and open scientific review.

Outcomes, which apparently are based on a combination of indicators, could result simply from the fact that, according to information described by CNN in a report on April 15, 2011, individuals are singled out for behaving arrogantly. Arrogant individuals stand a greater chance of being referred to a law enforcement official (LEO) than do those who not behave arrogantly. LEO referrals are related to 2 of the 4 the outcome measures (either by occurring individually or in combination with another indicator). Thus, almost by definition, the SPOT program has a higher probability of producing increases in outcome when compared with totally random selection. Positive SPOT outcomes are mostly due to observations that result in LEO interaction. These could be strongly related to things like "arrogant" behavior and be telling us little more than that, which is kind of a "duh?" result for such a serious investment of time and money. TAC had not been provided with enough information by the time of the April 6 hearing (when Mr. Willis indicated that the report had already been finalized) to determine significance and/or potential interaction with other variables. In summary, it is unclear what "effective" means in this context. The most significant outcomes in SPOT were related to LEO referrals. It is possible that the outcome of this program is no more than the observation that individuals who act like jerks might get arrested. What does that have to do with an effective, useful program?

Responses by Mr. Peter J. DiDomenica, Lieutenant Detective, Boston University Police

Questions submitted by Chairman Paul Broun

Q1. In your written testimony, you talk about your desire to see some sort of SPOT training provided for law enforcement personnel so that they can better coordinate and understand a situation when approached by a BDO who has suspicions about a traveler. Keeping in mind the limited resources we have in terms of federal dollars, can you expand on how critical such training would be? Would we be better off having fewer BDOs with more SPOT-trained LEOs?

A1. I believe that SPOT-trained police officers working in conjunction with the TSA are critical to the success of the SPOT program not only because of the ability of law enforcement to coordinate and understand the program but, most importantly, because of the absolute need for effective resolution of the suspicion. The BDOs are not empowered to detain, arrest, or deny access and lack law enforcement training and experience in questioning suspicious persons. Moreover, the BDOs do not have direct access to the criminal databases that law enforcement officers have access. The success of the program relies upon law enforcement officers (LEOs) who understand and use behavioral screening who follow through with denial of access, detention, or arrest when appropriate; otherwise, terrorists or other dangerous people will likely pass through the system because there will nothing obvious to justify denial of access or arrest such as a pre-existing arrest warrant or possession of contraband. The dilemma is that the most dangerous people, such as the 16 suspected terrorists who passed through SPOT airports, are generally not actively involved in a terrorist operation when boarding planes so that, short of finding an arrest warrant or contraband, there will be no basis for arrest. Even if they are operational and possess a weapon or explosive, there are still major gaps in weapon and explosive detection systems that present the significant risk of such weapon or explosive getting through the physical screening process. In my opinion it is absolutely critical that behavior assessment trained LEOs are present who are in a position to develop probable cause to arrest and who, absent such probable cause, are in a position to deny access when sufficient reasonable suspicion exists allowing the time for a more thorough investigation. Effective and reasonable security to prevent massive casualties from a terrorist attack on venues such as airports and mass transit significantly depends, in my opinion, upon behavior assessment trained LEOs who have the knowledge, ability, and confidence to deny access, in most cases temporarily, to such venues.

I believe the limited federal dollars available for SPOT screening would be better spent on training LEOs in behavior assessment and for providing federal support for overtime costs of deploying local and state LEOs for specific behavior assessment duties at airports. It seems to me that the American public will get "more bang for the buck" by enhancing the abilities of already trained and experienced law enforcement officers who can combine both the functions of being the "spotters" of suspicious behavior and being the "resolvers" of suspicious behavior. This would reduce the communication and understanding issues between TSA and LEOs that presently impede the success of the program. Moreover, the federal government would not be saddled with the costs of additional federal employees by contracting out the function to employees of state and local government. Such an approach would also reduce the civil liability exposure of the federal government as well. With this approach I believe there would be more effective prevention of terrorism with less expenditure of federal dollars.

Q2. I get the impression from your testimony that after the events of 9/11, particularly in light of your closeness to the situation, you felt the nation had to do something to prevent terrorism in the aviation sector. Your experience with Richard Reid appears to provide further evidence of that mentality.

a. Is that assessment of your mindset as you set about creating the program?

b. In the NRC's 2008 Report: Protecting Individual Privacy in the Struggle Against Terrorists - A Framework for Program Assessment, one of the conclusions reached by the 21-member Committee that published the report is:

> *In the aftermath of a disaster or terrorist incident, policy makers come under intense political pressure to respond with measures intended to prevent the event from occurring again. The policy impulse to do something (by which is usually meant something new) under these circumstances is understandable, but it is simply not true that doing something new is always better than doing nothing."*

b. How do you respond to that conclusion?

A2 (a.) I am not comfortable with the word "mentality" as used in the question as it implies, in my opinion, a certain rigidity and unwillingness to consider differing opinion perhaps to the point of being a zealot. I do not believe I had a "mentality" about having to do something to prevent terrorism construing the word "mentality" as I have explained. I did believe that our ability to screen passengers at airports was deficient and that it could be improved and that the Richard Reid example showed how reliance on physical screening without use of behavioral screening created a gap in security. I knew from my personal experience and from other police officers I worked with that persons who are engaged in dangerous or high risk activity tend to behave differently than persons not so engaged, particularly in the presence of a police officer or other official who could intercept them. I also learned through scientific literature that people's behavior changes when engaged in dangerous or high risk activity and that body language, mental state and paralinguistic attributes can be affected. It seemed reasonable to me then as it does today to use the ability of trained professionals to detect a person engaged in dangerous or high risk activity as another of layer of security at our airports provided the training was proper and the public's civil rights were protected through adhering to limitations on detentions and profiling based on the 4th Amendment and the Equal Protection Clause of the 14th Amendment. I do not believe I was under the impulse to do anything for the sake of doing anything but was motivated by addressing a gap in our security through reasonable, effective, and lawful means.

A2 (b-c.) I agree 100% with the danger presented by catastrophic events that can compel governments to respond without due deliberation and in haste sometimes with troubling and even devastating consequences. I have been an instructor in racial profiling and biased policing for over a decade and have included discussion of excesses by the government to respond to a serious incident or crisis. For example, the internment of more 100,000 Japanese Americans on the West Coast, mostly U.S. citizens, simply based on ancestry during World War II because of fears of an invasion or sabotage represents such an overreaction to a real threat. In fact, the U.S. Congress formally apologized to the survivors in 1988. The divisive issue of police racial profiling was spawned by overreaction to the real danger of drugs being transported on our highways. Well intentioned efforts to make communities safer resulted in those very communities feeling disenfranchised from law enforcement through the unlawful use of selective enforcement based on race. I was well aware of the danger to the American public from overreaction to the real threat of Islamic Extremist terrorism and made efforts to ensure our response was lawful and effective and consistent with our nation's values. I, like many security and law enforcement officials, found a gap in our aviation security and sought and found a means to address the gap, not because something had to be done but because something could be done. I would also like to point out that I was not a policy maker but a policy advisor and was not personally under any political pressure to do something. I was not an elected official nor did I directly serve elected officials. I could have simply carried out my duties as a police officer without having attempted to address the issue or passenger screening but chose to help because I felt I was the type of person who could balance the need for response to terrorism with the ability to do it effectively, lawfully, and ethically without undue haste and with proper deliberation.

Q3. Did you consult with any scientists before implementing the BASS program? What scientific literature did you research prior to the program?

a. Do you consider this review exhaustive or comprehensive?

b. Have you ever submitted the BASS system for outside review by Behavioral Scientists?

c. Did you encounter any criticisms– either through your research or by talking to people - about the validity of the BASS program?

A3. I consulted with co-panelist Dr. Paul Ekman and Dr. Mark Frank of the State University of New York at Buffalo. Then Massachusetts State Police Major Thomas Robbins and I went to Quantico, VA and spoke with the FBI Behavioral Sciences Unit (Eugene Ragala and Stephen Etter). We also spoke with Dr. Jessica Stern of the Harvard Kennedy School of Government.

Literature consulted included:

- Atran, Scott, University of Michigan, The Surprises of Suicide Terrorism, Discover Magazine, Vol. 24 No. 10 (October 2003)
- Lewis, Bernard, What Went Wrong

- The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks Upon the United States.
- Stern, Jessica, Harvard University John F. Kennedy School of Government, The Protean Enemy, Foreign Affairs, Volume 82 No. 4, July/August 2003, p. 27.
- Stern, Jessica, Terror in the Name of God
- Richardson, Louise, Harvard University professor, What Terrorists Want
- Pape, Robert, University of Chicago, Dying to Win, Database of every suicide attack from 1980 to 2003, 315 attacks
- Knapp, Mark, and Hall, Judith, Nonverbal Communication in Human Interaction
- Miller, Arthur G., editor, The Social Psychology of Good and Evil
- McDermott, Terry, Perfect Soldiers
- Grossman, Dave, On Killing, On Combat
- Dozier Jr., Rush, Why We Hate
- Barber, Benjamin, Jihad vs. McWorld
- Who Becomes a Terrorist and Why (US Government Report)
- Zimbardo, Phillip, Stanford Prison Experiment (1971)
- Milgram, Stanley, Obedience Experiments (1974)
- Givens, David B, Center for Nonverbal Studies, The Nonverbal Dictionary of Gestures, Signs & Body Language Cues (2003).
- Sageman, Marc, Former CIA caseworker and forensic psychologist, Study of 400 terrorists
- Meta-analysis on deception cues by Bella DePaulo, et al., 2003. Cues to Deception, Psychological Bulletin, 129(1):74-118, 2003
- Mehrabian, Albert, and Ferris, Susan R. "Inference of Attitudes from Nonverbal Communication in Two Channels," Journal of Consulting Psychology, Vol. 31, No. 3, June 1967, pp. 248-258
- Mehrabian, A. (1971). Silent messages, Wadsworth, California: Belmont
- Mehrabian, A. (1972). Nonverbal communication. Aldine-Atherton, Illinois: Chicago
- Facial expression of emotion; seven universal expressions of emotion. Ekman, Friesen, & O'Sullivan, 1988.
- Darwin, Charles, The Expression of Emotion in Man and Animals
- Testimony of Professor Jonathan Turley, Shapiro Professor of Public Interest, George Washington University Law School, before the U.S. House of Representatives Subcommittee on Aviation, February 27, 2002. Available on the Internet at http://www.house.gov/transportation/aviation/02-27-02/turley.html
- Ekman Ph.D., Paul, Telling Lies and Human Emotion Revealed

A3 (a.) I do not believe this review to be exhaustive but I do believe it was comprehensive.

A3 (b.) I asked Dr. Ekman, Dr. Frank, and the FBI Behavioral Sciences Unit to look at the program but this was not in the nature of a formal scientific review.

A3 (c.) I participated as a briefer for the JASON (Mitre Corporation) Summer Study "Badguyology" in June 2008 in which I presented information on BASS techniques. Their findings where that anecdotal evidence exists that police interviewing methodologies work at detecting deception and may be able to be validated and developed further. However, they also found that no scientific evidence exists to support the detection or inference of future behavior including intent. My discussions with Dr. Ekman, Dr. Frank and the FBI Behavioral Sciences Unit generally indicated the same assessment of BASS: that there was a general scientific foundation for changes in behavior related to persons engaged in high risk activity who did not want to be detected but specific studies would be needed to validate the use of specific behaviors and their significance.

Q4. What does the BASS/PASS training consist of? What behavior/cues/deviations did you look for?

A5. The following is the training outline of the BASS program showing all the components of the training:

INTRODUCTION

- War in the Homeland
- Policing in the Post 9/11 Environment
- Rationale for BASS
- What is BASS
- Is BASS Profiling?
- Benefits of BASS

BASS POLICY AND LEGAL CONSIDERATIONS

- Definitions
- Prohibition on Racial Profiling
- Voluntary Encounters

BASS GENERAL GUIDELINES AND PROCEDURES

- Methods of Contact
- Guidelines for Elevated and Reasonable Suspicion

UNDERSTANDING THE TERROR THREAT

- Islamic Fundamentalist Terror
- History of Conflict
- The Current Threat

STEP (1) OBSERVATION OF BEHAVIOR

- Theory of Behavioral Analysis
- Understanding Baselines
- Baseline Field Exercise
- Low Level Behavioral Indicators
- High Level Behavioral Indicators
- Surveillance Indicators
- Unusual Items in Baggage
- Explosive Components
- Suicide Bomber Indicators
- Detecting Bomb Activity in Vehicles and Buildings
- London Bombings
- 9/11 hijackers
- Evolving Suicide Bomber
- High and Low Risk Passengers

STEP (2) EXAMINATION OF TRAVEL DOCUMENTS

- Resident Alien
- Passport
- Visa
- I-94 and I-94W forms
- Elevated Suspicion Factors
- Terrorist Sponsoring and Terrorist Suspicious Countries

STEP (3) INTERVIEW

- Purpose of Interview
- Format of Questions o Travel/Visit Questions
  - Vehicle Stop Questions
  - Question Form and Technique
- Two-Step Baseline Approach to Resolving Elevated Suspicion
- Signs of Deception
- Analysis of Interview Videos
- Classroom Interview Exercise

STEP (4) RESOLUTION

- Three Dispositions of Person
- Case Studies

- *FIELD INTERVIEW EXERCISES COURSE CONCLUSION*

- Summary of Course
- Q & A
- Evaluations

The specific behavior/cues/deviations may be protected under TSA regulations as Sensitive Security Information so I cannot answer this question without further guidance from legal counsel.

Q5. Page two of Dr. Hartwig testimony states..How do you respond to Dr. Hartwig and Dr. Rubin's testimony?

A5. BASS is not a lie detection program: BASS is a program designed to detect behavioral changes associated with a person who is engaged in high risk or dangerous activity and to prevent such persons from entering critical infrastructure until the status of the person is resolved. Detection of deception constitutes one factor of many as part of an overall assessment of dangerousness and this factor, while useful, is not required for identification of potentially dangerous people. I have attended the following courses on interviewing that include detection of deception components

and this training indicates that with such interviewing training, police officers can improve their ability to detect deception:

Paul Ekman Group Training Division
Evaluating Truthfulness Train-the Trainer Workshop, February 16-18, 2006.

Institute of Analytic Interviewing
Interviewing, Credibility, and Emotion, January 10-14, 2005.

Department of the Treasury, Bureau of Alcohol, Tobacco, and Firearms
Analytic Interview School, April 19-23, 1999 at State Police New Braintree.

Wicklander - Zulawski & Associates
The Reid Method of Criminal Interviews and Interrogation, April 16-18, 1996 at State Police New Braintree.

Moreover, I am certified as a trainer in deception detection by the Paul Ekman Group Training Group and have conducted this training for the TSA and the Department of State. From my understanding of the research, there are techniques considered fairly reliable in detection of deception and that if used as part of an integrated approach that considers both emotional and cognitive aspects of deception and memory, the seriousness of the potential deception, alternative explanations for perceived cues, and evaluation of subject baseline, can allow police officers to be more effective and accurate in the assessment of credibility. I believe the DHS SPOT validation study provides striking evidence for the effectiveness of the SPOT/BASS techniques I designed: A high-risk traveler is nine times more likely to be identified using operational SPOT versus random screening and that this result was achieved by BDOs engaging 50,000 fewer passengers than the random selection process. When it came to arrests in this study, the SPOT program was found to be 50 times more effective than random screening. Moreover, the research by Dr. Frank cited in Dr. Ekman's testimony indicates that, "In a situation set up to resemble an airport security context, we could predict at 90% accuracy who intended to lie about an action which s/he had not yet taken. This was accomplished by analysis solely on their emotional reaction, eye contact, and nervous body behaviors. These are the types of actions security officers look for in behavioral observation programs. These results are the first study to show that intentions can be detected from behavior." Combining my training and experience and this recent research I am confident that properly trained LEOs have a significantly better than chance ability to detect potential terrorists and other dangerous people.

I agree with Dr. Rubin's testimony that shows there is an inclination by those who are involved in evaluations in the criminal and homeland/national security arena to be dismissive of scholarly research that may contradict their views. This is an aspect of basic human nature that we all tend to become defensive when our basic assumptions are challenged and this includes police officers, scientists, and congressmen. Nobody likes being told they are wrong. I have always tried to keep an open mind in my professional work and my work in developing SPOT/BASS was done in this way to the best of my ability. Most of what I learned and experienced pointed to the programs going in the right direction and I always welcomed review and advice. I welcome continued research and testing and know there is a great deal more to be learned. I agree with the GAO report 10-763 of May 2010 that called for more scientific validation of SPOT and I am personally disappointed that TSA did not do more to validate the program after I left in 2004. To be blunt in my opinion, TSA dropped the ball in its efforts to validate SPOT and, as a result have put many people and entities on the "spot" to defend it and to question it including myself, DHS, and this Subcommittee. But as Chairman Broun stated at the April 6, 2011 hearing, "The goal is not to throw out the proverbial baby with the bath water." I believe SPOT/BASS programs provide a critical layer in our multifaceted approach to aviation security and the effort to validate the programs, however belated, is worth our time and expense.

Thank you for this additional opportunity to address the Subcommittee.

Appendix II

Additional Materials Submitted for the Record

MATERIAL SUBMITTED BY MR. STEPHEN LORD, DIRECTOR, HOMELAND SECURITY AND JUSTICE ISSUES, GOVERNMENT ACCOUNTABILITY OFFICE

United States Government Accountability Office

GAO

Report to the Ranking Member, Committee on Transportation and Infrastructure, House of Representatives

May 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
Accountability * Integrity * Reliability

GAO-10-763

May 2010

GAO
Accountability · Integrity · Reliability

Highlights

Highlights of GAO-10-763, a report to the Ranking Member, Committee on Transportation and Infrastructure, House of Representatives

AVIATION SECURITY

Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges

Why GAO Did This Study

To enhance aviation security, the Transportation Security Administration (TSA) began initial testing in October 2003 of its Screening of Passengers by Observation Techniques (SPOT) program. Behavior Detection Officers (BDO) carry out SPOT's mission to identify persons who pose a risk to aviation security by focusing on behavioral and appearance indicators. GAO was asked to review the SPOT program. GAO analyzed (1) the extent to which TSA validated the SPOT program before deployment, (2) implementation challenges, and (3) the extent to which TSA measures SPOT's effect on aviation security. GAO analyzed TSA documents, such as strategic plans and operating procedures; interviewed agency personnel and subject matter experts; and visited 15 SPOT airports, among other things. Although the results from these visits are not generalizable, they provided insights into SPOT operations.

What GAO Recommends

GAO recommends that TSA, among other things, use an independent panel of experts to assist in validating SPOT, enhance SPOT data collection and analysis, fully utilize TSA resources to identify possible threats, and establish a plan to develop more outcome-oriented measures for SPOT. DHS reviewed a draft of this report and generally concurred with our recommendations although its plans do not fully address one of our recommendations.

View GAO-10-763 or key components. For more information, contact Stephen M. Lord at (202) 512-4379 or lords@gao.gov.

What GAO Found

Although the Department of Homeland Security (DHS) is in the process of validating some aspects of the SPOT program, TSA deployed SPOT nationwide without first validating the scientific basis for identifying suspicious passengers in an airport environment. A scientific consensus does not exist on whether behavior detection principles can be reliably used for counterterrorism purposes, according to the National Research Council of the National Academy of Sciences. According to TSA, no other large-scale security screening program based on behavioral indicators has ever been rigorously scientifically validated. DHS plans to review aspects of SPOT, such as whether the program is more effective at identifying threats than random screening. Nonetheless, DHS's current plan to assess SPOT is not designed to fully validate whether behavior detection can be used to reliably identify individuals in an airport environment who pose a security risk. For example, factors such as the length of time BDOs can observe passengers without becoming fatigued are not part of the plan and could provide additional information on the extent to which SPOT can be effectively implemented. Prior GAO work has found that independent expert review panels can provide comprehensive, objective reviews of complex issues. Use of such a panel to review DHS's methodology could help ensure a rigorous, scientific validation of SPOT, helping provide more assurance that SPOT is fulfilling its mission to strengthen aviation security.

TSA is experiencing implementation challenges, including not fully utilizing the resources it has available to systematically collect and analyze the information obtained by BDOs on passengers who may pose a threat to the aviation system. TSA's Transportation System Operations Center has the resources to investigate aviation threats but generally does not check all law enforcement and intelligence databases available to it to identify persons referred by BDOs. Utilizing existing resources would enhance TSA's ability to quickly verify passenger identity and could help TSA to more reliably "connect the dots." Further, most BDOs lack a mechanism to input data on suspicious passengers into a database used by TSA analysts and also lack a means to obtain information from the Transportation System Operations Center on a timely basis. TSA states that it is in the process of providing input capabilities, but does not have a time frame for when this will occur at all SPOT airports. Providing BDOs, or other TSA personnel, with these capabilities could help TSA "connect the dots" to identify potential threats.

Although TSA has some performance measures related to SPOT, it lacks outcome-oriented measures to evaluate the program's progress toward reaching its goals. Establishing a plan to develop these measures could better position TSA to determine if SPOT is contributing to TSA's strategic goals for aviation security. TSA is planning to enhance its evaluation capabilities in 2010 to more readily assess the program's effectiveness by conducting statistical analysis of data related to SPOT referrals to law enforcement and associated arrests.

Contents

Letter		1
	Background	8
	DHS Is Taking Action to Validate the Scientific Basis of TSA's SPOT Program but Opportunities Exist to Help Inform Future Program Decisions	14
	More Fully and Consistently Utilizing Available Information Technology Could Enhance TSA's Ability to Identify Threats to the Aviation System	31
	TSA Lacks Program Effectiveness Measures for SPOT but Is Taking Steps to Improve Evaluation Capabilities	39
	TSA Developed and Deployed SPOT Training but Further Action Could Enhance Its Effectiveness	49
	Conclusions	58
	Recommendations for Executive Action	60
	Agency Comments and Our Evaluation	62
Appendix I	**Scope and Methodology**	70
Appendix II	**DHS Comments**	77
Appendix III	**GAO Contacts and Staff Acknowledgments**	84
Tables		
	Table 1: Summary of Desirable Characteristics for Developing a Strategic Plan	28
	Table 2: Reasons for Arrests from SPOT Referrals, May 29, 2004 through August 31, 2008	44
	Table 3: SPOT Instructor Evaluation Ratings, 2006 to September 2008, and March 2009	54
	Table 4: TSA Training Standards and Evaluation Branch Recommendations for Improving SPOT Training and TSA Actions on the Recommendations	57
Figures		
	Figure 1: TSA's Layers of Aviation Security	9

Figure 2: The First Step in the SPOT Process: BDOs Observing Passengers About to Go Through Checkpoint Magnetometer 11
Figure 3: Budget and Personnel Growth in the SPOT Program, Fiscal Years 2007 through 2010 27
Figure 4: Passenger Boardings at SPOT Airports, May 29, 2004, through August 31, 2008 43

Abbreviations

AMRA	Aviation Modal Risk Assessment
BDO	Behavior Detection Officer
CBP	U.S. Customs and Border Protection
DEA	Drug Enforcement Agency
DHS	Department of Homeland Security
FAMS	Federal Air Marshal Service
FBI	Federal Bureau of Investigation
ICE	U.S. Immigration and Customs Enforcement
LEO	Law Enforcement Officer
NCIC	National Crime Information Center
NIPP	National Infrastructure Protection Plan
OMB	Office of Management and Budget
SOP	Standard Operating Procedures
SPOT	Screening of Passengers by Observation Techniques
S&T	Science and Technology Directorate
TSA	Transportation Security Administration
TSO	Transportation Security Officer

This is a work of the U.S. government and is not subject to copyright protection in the United States. The published product may be reproduced and distributed in its entirety without further permission from GAO. However, because this work may contain copyrighted images or other material, permission from the copyright holder may be necessary if you wish to reproduce this material separately.

United States Government Accountability Office
Washington, DC 20548

May 20, 2010

The Honorable John L. Mica
Ranking Member
Committee on Transportation and Infrastructure
House of Representatives

Dear Mr. Mica:

The terrorist attacks of September 11, 2001, highlighted the need to improve security within the nation's civil aviation system to deter persons seeking to repeat similar attacks on the nation's critical infrastructure. In October 2003, the Transportation Security Administration (TSA) of the Department of Homeland Security (DHS) conducted an operational test of the use of behavior detection techniques to screen passengers in an airport environment, and subsequently began training certain Transportation Security Officers (TSO)—TSA employees responsible for screening passengers and their property—in these techniques. These TSOs performed behavior observation as a collateral duty. Beginning in fiscal year 2007, TSA created separate Behavior Detection Officer (BDO) positions as part of the Screening of Passengers by Observation Techniques (SPOT) program.[1] According to TSA, the SPOT program is a derivative of other behavioral analysis programs that have been successfully employed by law enforcement and security personnel both in the United States and around the world, particularly that of Israel's airline, El Al.[2]

TSA designed SPOT to provide BDOs with a means of identifying persons who may pose a potential security risk at TSA-regulated airports[3] by focusing on behaviors and appearances that deviate from an established

[1]BDOs must have at least 12 months experience as a TSO, or related security work experience, and must pass a BDO training course.

[2]TSA cautions that the applicability of El Al's security processes to those used by TSA is constrained by differences in the scale of El Al's worldwide operations and the flexibilities that El Al has in implementing security processes compared to constraints on TSA. For example, El Al security screeners are encouraged to spend as much time with passengers as needed, and are not concerned whether passengers experience delays in boarding an aircraft.

[3]For the purposes of this report, the term "TSA-regulated airport" refers to a U.S. airport operating under a TSA-approved security program.

baseline, and that may be indicative of stress, fear, or deception. Passengers in an airport terminal, including those waiting in security checkpoint lines, are observed by the BDOs to determine if their behavioral and appearance indicators—which are assigned varying points by SPOT—have (in combination) exceeded a predetermined numerical threshold. In cases where the passenger exceeds the threshold, the passenger is referred for additional screening by BDOs and a TSO. During this referral screening, if the passenger exhibits behaviors that exceed another numerical threshold, they are to be referred to a law enforcement officer (LEO) for further investigation. In addition to observing passengers at airport checkpoints, BDOs may patrol throughout an airport terminal, and sometimes participate in other activities, such as TSA's Visible Intermodal Prevention and Response team operations. These teams are responsible for periodically augmenting security at air and ground transportation facilities around the country.[4]

As of March 2010, TSA deployed about 3,000 BDOs at an annual cost of about $212 million; this force increased almost fifteen-fold between March 2007 and July 2009. BDOs have been selectively deployed to 161 of the 457 TSA-regulated airports in the United States at which passengers and their property are subject to TSA-mandated screening procedures.[5] The conference report accompanying the fiscal year 2010 DHS appropriations act provided that $211.9 million of aviation security funding was for the SPOT program.[6] The administration has requested $232 million for SPOT for fiscal year 2011, a $20.2 million (9.5 percent) increase over the current funding level. This increase would support a workforce increase from about 3,000 to 3,350 BDOs. If this funding request is approved and maintained, SPOT would cost about $1.2 billion over the next 5 years.

[4]Visible Intermodal Prevention and Response teams are comprised of federal air marshals, surface transportation security inspectors, TSOs, BDOs, and canines.

[5]TSA classifies its regulated airports in the United States into one of five categories—X, I, II, III, and IV. Generally, category X airports have the largest number of passenger boardings and category IV airports have the least.

[6]See H.R. Rep. No. 111-298 at 77 (2009) (Conf. Rep.). The conference report directed TSA to report, no later than 60 days after enactment, on the scientific basis for using behavior pattern recognition for observing airline passengers for signs of hostile intent, the effectiveness of the SPOT program in meeting its goals and objectives, and the justification for expanding the program. The conference report also directed us to review this report and to provide our findings to the Committees no later than 120 days after the TSA report is submitted. TSA completed its report to Congress on March 15, 2010.

You asked us to address SPOT's development and implementation. This report addresses the following questions:

1. To what extent did TSA determine whether SPOT had a scientifically validated basis for identifying passengers before deploying it and utilize recognized best practices during SPOT's development?
2. What management challenges, if any, have emerged during the implementation of SPOT at the nation's airports?
3. To what extent has TSA measured SPOT's effect on aviation security?
4. To what extent has TSA incorporated the attributes of an effective training program into the training for SPOT?

This report is a public version of the restricted report (GAO-10–157SU) that we provided to you on May 14, 2010. DHS and TSA deemed some of the information in the restricted report as sensitive security information, which must be protected from public disclosure. Therefore, this report omits this information. Although the information provided in this report is more limited in scope, it addresses the same questions as the restricted report. Also, the overall methodology used for both reports is the same.

To determine the extent to which TSA determined whether SPOT had a scientifically validated basis for identifying passengers who may pose a risk to aviation security before deploying it, we reviewed literature on behavior analysis by subject matter experts, and analyzed relevant reports and books on the topic. These included a 2008 study by the National Research Council of the National Academy of Sciences that included a discussion section on deception and behavioral surveillance, as well as other issues related to behavioral analysis.[7] We interviewed seven recognized experts in the field, and an expert on emergency responses to terror attacks and mathematical models in operations management.[8] Although the views of these experts cannot be generalized across all experts on behavior analysis, because we selected these individuals based on their publications on behavioral analysis or related topics, their recognized accomplishments and expertise, and, in some cases, TSA's use of their work or expertise to design and review the SPOT program's

[7]National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008). We reviewed the approach used and the information provided in this study and found the study and its results to be reliable for the purposes for which we used it in this report.

[8]See app. I for additional information on the experts we interviewed.

behaviors, they provided us with an understanding of the fundamentals of behavior analysis, and its use in airports. We also interviewed cognizant officials from other U.S. government agencies that utilize behavior analysis in their work, including U.S. Customs and Border Protection (CBP), the U.S. Secret Service, the Federal Air Marshall Service (FAMS), and the Federal Bureau of Investigation (FBI).[9] To better understand how SPOT incorporated expertise on behavior analysis for aviation security, we also interviewed current and retired officials of Israel's El Al Airlines, whose security processes TSA cites as providing part of the basis of the SPOT program.[10]

To determine to what extent TSA utilized best practices during SPOT's development—including carrying out a comprehensive risk assessment, a cost-benefit analysis, and a strategic plan—we interviewed program officials and reviewed related program documentation, including briefings used in the course of developing and fielding SPOT, strategic plans, and standard operating procedures.[11] We compared these documents to DHS's 2006 *Cost Benefit Analysis Guidebook*,[12] Office of Management and

[9]For reasons of scope, we did not assess the scientific basis of the methods and processes used by these agencies in their application of behavioral detection.

[10]Although SPOT is based in some respects on El Al's aviation security program, El Al's processes differ in substantive ways from those used by the SPOT program. In particular, El Al does not use a list of specific behaviors with numerical values for each, or a numerical threshold to determine whether or not to question a passenger; rather, El Al security officers utilize behavioral indicators as a basis for interviewing all passengers boarding El Al passenger aircraft, and accessing relevant intelligence databases, when deemed appropriate. In addition, El Al officials told us that they train all their personnel—not just security officers—in elements of behavior analysis, and conduct covert tests of their employees' attentiveness at frequent intervals. According to these officials, El Al also permits what is termed "profiling," in which passengers may be singled out for further questioning based on their nationality, ethnicity, religion, appearance, or other ascriptive characteristics, but these are not the only basis on which a passenger may be questioned. In addition, El Al security officers are empowered to bar any passenger from boarding an aircraft. The scale of El Al operations is considerably smaller than that of major airlines operating within the United States. As of 2008, El Al had a fleet of 34 aircraft. In Israel, El Al operates out of one hub airport, Ben-Gurion International, and also flies to Eilat, a city in southern Israel; in contrast, there are 457 TSA-regulated airports in the United States. In 2008, El Al had passenger boardings of about 3.6 million; in contrast, Southwest Airlines alone flew about 102 million passengers in the same year.

[11]Unless otherwise noted in the report, we refer to the SPOT strategic plan issued in March 2007.

[12]DHS, *Cost Benefit Analysis Guidebook* (Washington, D.C.: Feb. 1, 2006).

Budget (OMB) guidance,[13] and DHS's 2006 and 2009 National Infrastructure Protection Plans (NIPP), which set forth a risk management framework to guide security decision making and resource allocation decisions, and our previous work on the characteristics of an effective strategic plan.

To identify any challenges that emerged during implementation of the SPOT program, we conducted field site visits to 15 TSA-regulated airports with SPOT that represent almost 10 percent of the 161 TSA-regulated airports with SPOT to observe operations and meet with key program personnel.[14] We chose airports with high, medium, and low passenger volume; airports with BDOs who are TSA (i.e., government) employees and an airport with BDOs employed by contractors as part of the TSA Screening Partnership Program; and airports with LEOs who were identified by TSA as having received some form of behavior detection training and airports where they were not known to have received such training.[15] We also selected airports on the basis of TSA's assessment of which ones are at highest risk of attack by terrorists, including the 2 that ranked the highest, as reported in TSA's Current Airport Threat Assessment.[16] Since the airports we selected range broadly in terms of passenger volume, physical size and layout, geographic location, and potential value as a target for terrorism, among other things, the results from these visits are not generalizable to other airports. However, these visits provided helpful insights into the operation of SPOT at airports. In addition, to determine whether challenges emerged in implementing SPOT, we compared TSA's approach for implementing and managing

[13]OMB, Circular No. A-94, *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs* (Washington, D.C.: October 1992); Circular No. A-4, *Regulatory Analysis* (Washington, D.C.: Sept. 2003).

[14]See app. I for additional details on the airports we visited.

[15]At airports participating in TSA's Screening Partnership Program, private-sector contractors perform screening activities, including SPOT, in accordance with TSA requirements and oversight. See 49 U.S.C. § 44920. Unless otherwise specified, references to TSOs include private-sector contract screeners. For more information, see GAO, *Aviation Security: Progress Made to Set Up Program Using Private-Sector Airport Screeners, but More Work Remains*, GAO-06-166 (Washington, D.C.: Mar. 31, 2006).

[16]The TSA Current Airport Threat Assessment is a threat estimate designed to provide a snapshot of the current terrorist threat to airports in the United States as well as for major international airports serving as last points of departure for U.S. airlines.

SPOT to our *Standards for Internal Control in the Federal Government*[17] and to risk management principles we had previously identified.[18] In reviewing TSA's approach to developing and implementing SPOT, we considered relevant laws, regulations, and other materials, including those related to privacy, such as TSA's Privacy Impact Assessments. To obtain comparative data on how SPOT had been implemented at different airports across the nation, we conducted a survey of all Federal Security Directors responsible for security operations at TSA-regulated airports with SPOT.[19] (This accounted for all 161 TSA-regulated airports with SPOT because a single Federal Security Director may be responsible for several airports.) We obtained a 100 percent response rate. This survey asked, among other things, about the relationship between LEOs and the airport authority and BDOs. In addition, to understand the interaction of BDOs and LEOs, as well as other SPOT implementation issues, at each of the 15 TSA-regulated airports we visited we spoke with BDO managers, Federal Security Directors, Assistant Federal Security Directors, 1 or 2 BDOs, and 1 or 2 LEOs.

To determine the extent to which TSA has measured SPOT's effect on aviation security, we obtained and analyzed the TSA SPOT referral database,[20] which aims to record all incidents in which passengers who have passed through the checkpoint are sent to SPOT referral screening for additional questioning and screening of property and person. The database also maintains records of instances where passengers were referred by a BDO to a LEO for questioning. We assessed the reliability of the SPOT referral data by (1) performing electronic testing of required data elements, (2) reviewing existing information about the data and the

[17]GAO, *Standards for Internal Controls in the Federal Government*, GAO/AIMD-00-21.3.1 (Washington, D.C.: November 1999).

[18]See GAO, *Risk Management: Further Refinements Needed to Assess Risks and Prioritize Protective Measures at Ports and Other Critical Infrastructure*, GAO-06-91 (Washington, D.C.: Dec. 15, 2005) and *Transportation Security: Comprehensive Risk Assessments and Stronger Internal Controls Needed to Help Inform TSA Resource Allocation*, GAO-09-492 (Washington, D.C.: Mar. 27, 2009).

[19]Federal Security Directors are the highest ranking TSA officials responsible for security operations at TSA-regulated airports. See 49 U.S.C. § 44933. They and their assistants coordinate with both federal and nonfederal entities present at their airports, including the FAMS, the Drug Enforcement Administration, and CBP. When appropriate, Federal Security Directors may bar an individual from boarding an aircraft.

[20]The SPOT referral data we analyzed covered the period May 29, 2004, through August 31, 2008. These were the data available at the time of our analysis.

system that produced them, and (3) interviewing agency officials knowledgeable about the data. We found a number of problems related to how the data were collected and recorded that are discussed later in this report. As a result, we were unable to use the SPOT referral data to assess whether any behavior or combination of SPOT behaviors could be used to reliably predict the final outcome of an incident involving the use of SPOT. However, with the stated limitations in mind, and after resolving certain contradictions and anomalies in the database, we utilized the SPOT referral data to provide examples of information used by TSA to report on the program's performance, including a count of arrests and the reasons for those arrests. In addition, to determine if individuals who were later charged with or pleaded guilty to terrorism-related offenses had transited SPOT airports and whether TSA could obtain information from these transits to enhance its understanding of terrorist behaviors, we reviewed CBP and Department of Justice information to (1) identify individuals who were charged with or pleaded guilty to terrorism-related offenses and (2) determine if these individuals had, prior to being charged, transited airports where SPOT had been deployed. Further, we used our survey of Federal Security Directors at SPOT airports to determine the extent to which video surveillance cameras, which could make video recordings of terrorists transiting airports, are present at checkpoints.

To assess the extent that SPOT training incorporates the attributes of an effective training program, we had TSA training experts complete a training assessment tool that we developed using guidance we prepared in our previous work for assessing training courses and curricula.[21] To better understand how other entities train their employees in behavior detection, and what their curricula include, we conducted site visits to the Secret Service, CBP, FAMS, and the FBI, and also interviewed nongovernmental experts on aspects of behavior detection training. We interviewed BDOs and BDO managers about the SPOT training. In addition, we interviewed El Al officials with regard to how El Al trains and tests its personnel in behavior recognition and analysis. Appendix I contains additional details about our scope and methodology.

We conducted this performance audit from May 2008 through May 2010 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain

[21]GAO, *Human Capital: A Guide for Assessing Strategic Training and Development Efforts in the Federal Government*, GAO-04-546G (Washington, D.C.: Mar. 1, 2004).

sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

Background

The Aviation and Transportation Security Act established TSA as the federal agency with primary responsibility for securing the nation's civil aviation system, which includes the screening of all passenger and property transported by commercial passenger aircraft.[22] TSA currently has direct responsibility for, or oversees the performance of, security operations at approximately 457 TSA-regulated airports in the United States implementing security requirements in accordance with TSA-approved security programs and other TSA direction.[23] At TSA-regulated airports, prior to boarding an aircraft, all passengers, their accessible property, and their checked baggage are screened pursuant to TSA-established procedures, which include, for example, passengers passing through security checkpoints where they and their identification documents are checked by TSOs and Travel Document Checkers, or by Screening Partnership Program employees.

TSA uses multiple layers of security to deter, detect, and disrupt persons posing a potential risk to aviation security. These layers include three principal types of screening employees at airport checkpoints—Travel Document Checkers, who examine tickets, passports, and other forms of identification; TSOs, who examine property, including checked baggage, and persons using x-ray equipment and magnetometers, as well as other devices; and BDOs, using SPOT to assess passenger behaviors and appearance.[24] BDOs are the only type of TSA screening employees not deployed to all TSA-regulated airports and all checkpoints within the

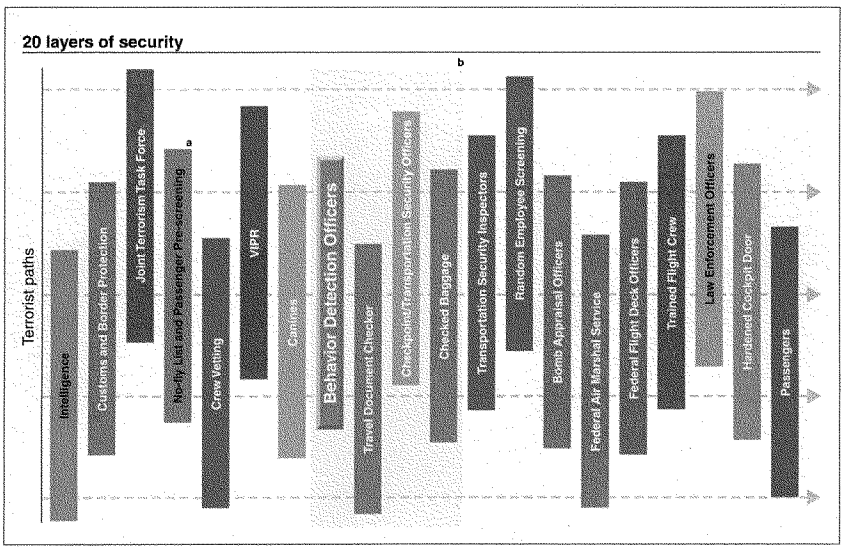
[22]See Pub. L. No. 107-71, 115 Stat. 597 (2001). For purposes of this report, "commercial aircraft" refers to a U.S. or foreign-based air carrier operating under TSA-approved security programs with regularly scheduled passenger operations to or from a U.S. airport.

[23]See 49 C.F.R. pt. 1542. Some commercial airports with fewer than 2,500 annual enplanements (passengers boarding an aircraft) do not have TSA-approved screening processes. Enplanements are the number of paying passengers on a scheduled or nonscheduled (charter) flight. Infants and airline personnel are not included. A stop at an airport is not considered an enplanement if the passenger does not transfer aircraft.

[24]Private-sector screeners under contract to and overseen by TSA, and not TSOs, perform screening activities at airports participating in TSA's Screening Partnership Program. See 49 U.S.C. § 44920.

airports where it is deployed on a regular basis. TSA deployed SPOT as an added layer of security to help deter terrorists attempting to exploit TSA's focus on prohibited items and other potential security weaknesses. Other security layers cited by TSA include intelligence gathering and analysis; passenger prescreening; random canine team searches at airports; federal air marshals; reinforced cockpit doors; federal flight deck officers; the passengers themselves; as well as other measures both visible and invisible to the public. Figure 1 shows TSA's 20 aviation security layers.

Figure 1: TSA's Layers of Aviation Security



Source: TSA.

[a] The No-Fly List is used to identify individuals who should be prevented from boarding an aircraft; it contains applicable records from the FBI's Terrorist Screening Center consolidated database of known or suspected terrorists.

[b] The four layers inside the grey bar are screening layers of security applied to passengers and their property.

The grey area in figure 1 highlights four layers that apply to passengers and their property as they seek to board an aircraft. Airport LEOs, another layer of security cited by TSA, do not report to TSA and may not maintain a physical presence at smaller TSA-regulated airports. According to TSA, each one of these layers alone is capable of stopping a terrorist attack. In combination, TSA states that their security value is multiplied, creating a much stronger system, and that a terrorist who has to overcome multiple security layers in order to carry out an attack is more likely to be preempted, deterred, or to fail during the attempt.

SPOT Uses Behavior Detection Techniques to Assess Passenger Behaviors and Appearances

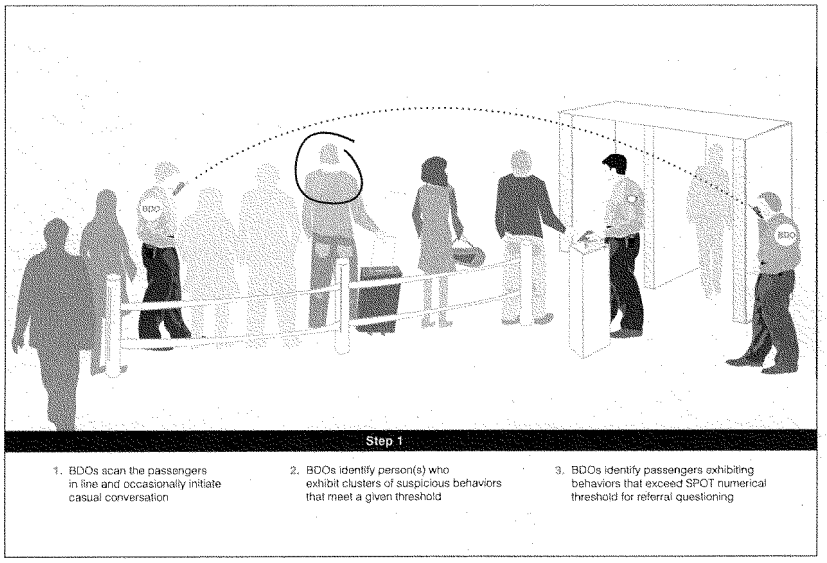
The SPOT program utilizes behavior observation and analysis techniques to identify potentially high-risk passengers. Individuals who exhibit suspicious behaviors, including both physical and appearance indicators, may be required to undergo additional screening. Field agents and law enforcement officers of other federal agencies and entities—such as the FBI, the Secret Service, CBP, and FAMS—utilize elements of behavior detection analysis as a part of their work. In addition, some foreign entities, such as Israel's El Al airlines, use behavior detection and analysis techniques as part of their security efforts. However, TSA emphasized to us that the SPOT program is unique among these entities because it uses a point system to help identify suspicious persons on the basis of their behavior and appearance and because behavior detection and analysis are the central focus of SPOT. Officials from the other agencies stated that their field personnel incorporate behavior detection as one of many skills used in their work; in contrast, behavior detection is the primary element of the BDOs' work.

SPOT trains BDOs to look for and recognize facial expressions, body language, and appearance that indicate the possibility that an individual is engaged in some form of deception and fears discovery. These behaviors and appearances are listed on a SPOT score sheet used in SPOT training.

Passenger behavior and appearance are to be compared by the BDOs—who typically work in two-person teams. BDOs are expected to "walk the line"—that is, to initiate casual conversations with passengers waiting in line, particularly if their observations led them to question someone exhibiting behaviors or appearances on the SPOT checklist. As the BDOs walk the line, and the passenger with SPOT indicators is reached, a casual conversation is used to determine if there is a basis for observed behaviors or appearances on the checklist. In most instances, these conversations provide information to the BDOs that permits them to consider the issue resolved, and hence not a security concern. Figure 2 below illustrates the

first step of the three-step SPOT process, the BDO-passenger interaction at a checkpoint prior to the passenger passing through a magnetometer.

Figure 2: The First Step in the SPOT Process: BDOs Observing Passengers About to Go Through Checkpoint Magnetometer



Sources: GAO (analysis), ArtExplosion (clip art), TSA (data).

Note: Circle around passenger shows a person who is exhibiting a cluster of suspicious behaviors.

As shown in figure 2, passenger behavior and appearance are observed by the BDOs as passengers wait in line for screening at a security checkpoint. Even if the checkpoint is busy, the BDOs must attempt to visually scan all the passengers waiting in line, as well as persons near the checkpoint, to determine if any are showing behaviors or appearances on the SPOT checklist. According to TSA, on average a BDO has approximately 30 seconds to assess each passenger while the passenger waits in line. For passengers exhibiting indicators above baseline conditions, the BDOs are to (mentally) add up the points assigned to each indicator they observe. Both BDO team members must agree that observed indicators have exceeded the predetermined numerical threshold, although they do not have to identify the same indicators the passenger exhibited. In instances when a passenger's SPOT indicators place them above the numerical threshold, and the passenger has placed their property on the conveyor belt for x-raying, and has walked through the magnetometer or equivalent screening device for passengers, he or she will be directed to the second step of SPOT, referral screening. This involves additional questioning and physical search of their person and property by BDOs and TSOs. This referral screening occurs in the checkpoint area.

If the passenger's behavior escalates further—accumulating more points based on the SPOT checklist—the BDOs are to refer the passenger to a LEO. A referral to a LEO is a potential third step in the SPOT process. BDOs are not LEOs—they do not conduct criminal investigations, carry weapons, or make arrests.

After a passenger has been referred by the BDOs to a LEO, the LEO is then expected to independently determine, through additional investigation, such as questioning the passenger and, if appropriate, by conducting an identity verification and background check through the FBI's National Crime Information Center (NCIC), whether sufficient grounds exist to take further action, such as detaining or arresting the passenger. TSA officials who are LEOs also have access to NCIC, such as an airport's Assistant Federal Security Director for Law Enforcement or federal air marshals. NCIC is the FBI's computerized index of criminal justice information (i.e., criminal record history information, fugitives, stolen properties, and missing persons), available to federal, state, and local law enforcement

and other criminal justice agencies at all times.[25] Similarly, other federal LEOs also have such access, including CBP, and Drug Enforcement Agency (DEA) personnel. However, since both local and federal LEOs have other responsibilities, and may not be present at each operating checkpoint, BDOs may have to seek them out to request an NCIC check. According to TSA, aside from requiring that an airport maintain a law enforcement presence,[26] it exercises no jurisdiction over the law enforcement activities of non-TSA officers or entities at an airport; thus, it cannot require LEOs to conduct an NCIC check or to provide BDOs with information about the ultimate disposition of cases referred by them to LEOs.

Once the LEO concludes his or her investigation and determines whether the passenger will be arrested or detained, TSA officials are to evaluate the security concerns to determine whether to allow the passenger to proceed to the boarding gate. (In some instances, a LEO might choose not to arrest or detain a passenger; TSA would then decide whether the infraction was sufficiently serious to necessitate barring the passenger from boarding.) After a referral incident has been resolved, BDOs are to enter information about the incident into TSA's SPOT referral database. The data entered are to include time, date, location of the incident, behaviors witnessed, prohibited items found (if any), and information on the LEO's response (if applicable), such as whether the LEO questioned the passenger, arrested the individual, or released the passenger. The SPOT referral database contains no personal identifying information about passengers.

SPOT Has Been Deployed in Phases

The SPOT program began with pilot tests in 2003 and 2004 at several New England airports, in which TSA began using uniformed BDOs at airport checkpoints. After some initial pilot projects and test deployments, 644

[25]These requests would typically be made to the law enforcement entity employing the LEO, such as the airport authority police department. The department would have a computer that can access NCIC. According to the FBI's website, the NCIC database consists of 19 files or databases. Seven are property files which contain records for articles, boats, guns, license plates, securities, vehicles, and vehicle and boat parts. Twelve are person files are the Convicted Sexual Offender Registry, Foreign Fugitive, Identity Theft, Immigration Violator, Missing Person, Protection Order, Supervised Release, Unidentified Person, U.S. Secret Service Protective, Violent Gang and Terrorist Organization, and Wanted Person Files. The Interstate Identification Index, which contains automated criminal history record information, is also accessible through the same network as NCIC. The Violent Gang and Terrorist Organization file includes the names of known or suspected terrorists.

[26]See 49 CFR §§ 1542.215, .217.

BDOs were deployed to 42 airports in the first phase of the program from November 2006 through June 2007. As of March 2010, about 3,000 BDOs utilizing SPOT were deployed at 161 of 457 TSA-regulated airports.[27]

BDO eligibility is restricted to TSOs with at least 12 months of TSO experience, or others with related security experience. Applicants must apply and be accepted into the BDO training program. The training includes 4 days of classroom courses, followed by 3 days of on-the-job training. BDOs must memorize all of the behaviors and appearances on the SPOT checklist, as well as the point value assigned to each, in order to be able to add these up to determine if a passenger should be sent to SPOT referral screening. BDO applicants must also pass a job knowledge test at the conclusion of the training. The test includes related multiple choice questions, true or false statements, and case-based scenarios.

DHS Is Taking Action to Validate the Scientific Basis of TSA's SPOT Program but Opportunities Exist to Help Inform Future Program Decisions

Although DHS is in the process of validating the way in which the SPOT program utilizes the science of behavior detection in an airport environment, TSA deployed SPOT nationwide before first determining whether there was a scientifically valid basis for using behavior and appearance indicators as a means for reliably identifying passengers as potential threats in airports. TSA reported that it deployed SPOT before a scientific validation of the program was completed in response to the need to address potential threats to the aviation system that would not necessarily be detected by existing layers of aviation security. TSA stated that no other large-scale U.S. or international screening program incorporating behavior- and appearance-based indicators has ever been rigorously scientifically validated. While TSA deployed SPOT on the basis of some risk-related factors, such as threat information and airport passenger volume, it did not use a comprehensive risk assessment to guide its strategy of selectively deploying SPOT to 161 of the nation's 457 TSA-regulated airports. TSA also expanded the SPOT program over the last 3 years without the benefit of a cost-benefit analysis of SPOT. Additionally, TSA's strategic plan for SPOT could be improved by the inclusion of desirable characteristics identified in our prior work, such as risk assessment information, cost and resources analysis, and a means for collaboration with other key entities.

[27]TSA-regulated airports have regular commercial passenger service and comply with TSA regulations for passengers and their property in order to operate.

TSA Is in the Process of Validating the Scientific Basis Used to Identify Passengers with SPOT Behaviors

TSA proceeded with deploying SPOT on a nationwide basis before determining whether the list of passenger behaviors and appearances underpinning the SPOT program were scientifically validated, and whether these techniques could be applied for counterterrorism purposes in an airport environment. In 2008, a report issued by the National Research Council of the National Academy of Sciences noted that behavior and appearances monitoring might be able to play a useful role in counterterrorism efforts but stated that a scientific consensus does not exist regarding whether any behavioral surveillance or physiological monitoring techniques are ready for use in the counterterrorist context given the present state of the science.[28] The report also stated that the scientific evidence for behavioral monitoring is preliminary in nature.[29] According to the report, an information-based program, such as a behavior detection program, 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 documentation on the program's effectiveness should be reviewed by an independent entity capable of evaluating the supporting scientific evidence. The report also stated that often scientists and other experts can help independently assess the scientific evidence on the effectiveness of a program. A contributor to the National Research Council report also stated that no conclusive research has been conducted to determine if behavior detection can be reliably used on a larger scale, such as in an airport setting, to identify persons intending to cause harm to the aviation system.

While TSA and DHS's Science and Technology (S&T) Directorate officials agreed that SPOT was deployed before its scientific underpinnings were fully validated, they stated that no large-scale U.S. or international operational screening program incorporating behavior- and appearance-

[28]National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008). The report's preparation was overseen by the National Academy of Sciences Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals. Although the report addresses broader issues related to privacy and data mining, a senior National Research Council official stated that the committee included behavior detection as a focus because any behavior detection program could have privacy implications.

[29]Specifically, the report states that the scientific support for linkages between behavioral and physiological markers and mental state is strongest for elementary states, such as simple emotions; weak for more complex states, such as deception; and nonexistent for highly complex states, such as when individuals hold terrorist intent and beliefs.

based indicators has been rigorously scientifically validated. These officials also questioned the findings of the National Research Council report and stated that the study lacked sufficient information for its conclusions because it did not consider recent findings from unpublished DHS, defense, and intelligence community studies.[30] However, National Research Council officials stated that an agency should be cautious about relying on the results of unpublished research that has not been peer reviewed, such as that generated by DHS and the defense and intelligence community, and using unpublished work as a basis for proceeding with a process, method, or program.[31] Moreover, we have previously reported that peer review is widely accepted as an important quality control mechanism that helps prevent the dissemination of potentially erroneous information.[32]

In addition to the unpublished research, TSA told us that the SPOT program was based on operational best practices from law enforcement, defense, and the intelligence communities. According to TSA officials, the agency based its choice of SPOT behavior, appearance, and deception indicators on existing research and training programs. For example, TSA cited research on emotions and their behavior indicators by Dr. Paul Ekman,[33] interviewing and interrogation by Stan Walters,[34] and nonverbal

[30]DHS's S&T Directorate could not provide us with specific contacts related to the sources of this research.

[31]Peer review is the process of subjecting an author's scholarly work, research, or ideas to the scrutiny of others who are experts in the same field. Such review is considered a form of scientific validation.

[32]For example, we reported that the National Institutes of Health did not post its researchers' final reports because the risks associated with posting results that have not been scrutinized and validated by peer review are too great. See GAO, *University Research: Most Federal Agencies Need to Better Protect against Financial Conflicts of Interest*, GAO-04-31 (Washington, D.C.: November 2003).

[33]Dr. Ekman is professor emeritus of psychology at the University of California Medical School, San Francisco, and is considered one of the world's foremost experts on facial expressions. His books include: *Emotions Revealed: Recognizing Faces and Feelings to Improve Communications and Emotional Life* (New York: Holt and Company, 2003); *Emotion in the Human Face* (New York: Pergamon Press, 1972); *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues* (Englewood Cliffs, N.J.: Prentice-Hall, 1975). Dr. Ekman has published more than 100 articles.

[34]Mr. Walters is the author of the *Principles of Kinesic Interview and Interrogation: 2nd Edition* as well as numerous training materials related to interviewing and interrogation techniques.

indicators by Dr. David Givens[36] and Dr. Mark Frank[36] as support for the choice of several of the behavior indicators. According to TSA, its development of the SPOT program was based on related DHS research and information from the training curricula of other federal agencies, such as the Federal Transit Administration and the Bureau of Alcohol, Tobacco, Firearms, and Explosives.[37]

As with the SPOT behavior indicators, TSA told us that it sought input in creating the SPOT point scoring system from subject matter experts and from participants in TSA's SPOT working group, which consisted of law enforcement officials from agencies such as FBI, DEA, and local law enforcement officials.[38] While TSA officials said that they coordinated with relevant subject matter experts, such as Dr. Ekman, and based the SPOT scoring system on existing research and training programs, no validation of the behavior, appearance, and deception indicators was conducted prior to the deployment of SPOT in November 2006. According to TSA officials, they used professional judgment in developing the SPOT

[36]Dr. Givens is the director of the nonprofit Center for Nonverbal Studies, in Spokane, Washington. Dr. Givens is the author of *Love Signals: A Practical Field Guide to the Body Language of Courtship* (St. Martin's, New York, 2005) and *Crime Signals: How to Spot a Criminal Before You Become a Victim* (St. Martin's, 2008). The Center's Web site links to Dr. Givens' online reference tool, *The Nonverbal Dictionary of Gestures, Signs and Body Language Cues*. Dr. Givens said that he had did not know which nonverbal indicators had been selected by TSA for use in SPOT, that he had not been asked by TSA to review their choices from his list, or to review other aspects of the SPOT program. According to Dr. Givens, attempting to detect more than nine nonverbal indicators would be difficult for most individuals, even those trained in behavior detection.

[36]Dr. Frank is Associate Professor, Department of Communication, College of Arts and Sciences, at the University at Buffalo, State University of New York. He is on the Advisory Board of the University's Center for Unified Biometrics and Sensors, and has conducted research supported by DHS, the Defense Advanced Research Projects Agency, and the National Science Foundation. Dr. Frank told us that he had observed SPOT at an airport and had some coordination with TSA. However, he said that he had not reviewed the SPOT training curriculum or the SPOT scoring system. Dr. Frank stated that no study has been performed to validate use of behavior detection in an airport setting.

[37]According to DHS's S&T Directorate, it completed a study on suicide bomber indicators in July 2009. The program manager stated that they reviewed 157 documents and identified approximately 1,200 suicide indicators, which were similar to SPOT suicide bomber indicators. S&T stated that the study provides preliminary support for the detection of suicide bomber indicators and that SPOT represents best practices from defense and intelligence organizations.

[38]According to TSA, the FBI participated in discussions related to SPOT's development in 2006.

point system and stated that the purpose of developing the scoring system was to increase the objectivity of the SPOT process.

Dr. Ekman stated that, in his opinion, and after reviewing the scoring system and observing the program in operation, it was not clear whether the SPOT behaviors and appearances, and the related point system, could be used effectively in an airport environment because no credible validation research on this issue had been conducted. He noted, for example, that research is needed to identify how many BDOs are required to observe a given number of passengers moving at a given rate per day in an airport environment, or the length of time that such observation can be conducted before observation fatigue affects the effectiveness of the personnel. He commented that observation fatigue is a well-known phenomenon among workers whose work involves intense observation, and that it is essential to determine the duration of effective observation and to ensure consistency and reliability among the personnel carrying out the observations.

DHS has recognized the need to conduct additional research to scientifically validate the use of the SPOT behavioral indicators in an airport environment. DHS's S&T Directorate began research in 2007 to determine if there is a statistically significant correlation between the SPOT behaviors exhibited by airport passengers and finding airport passengers with prohibited items (such as weapons), false documents, and illegal drugs or who pose a potential risk to aviation security. According to S&T, this research is expected to be completed in fiscal year 2011 and is to include three key elements. First, the study's purpose is to assess the reliability of the SPOT program by analyzing TSA's SPOT database to determine patterns of BDO scoring to measure consistency across BDOs, teams, locations, and other variables. Second, the study aims to compare the current implementation of SPOT to random passenger screening. Specifically, according to S&T officials, 130,000 passengers are to be randomly selected for additional SPOT referral screening. The study's design states that data collected about these passengers will be compared to data for passengers screened through the normal SPOT process. S&T officials expect that the results of this element of the study will provide a better understanding of how SPOT compares to random selection, as well as providing a baseline of each indicator present in the traveling public. Third, the study also aims to utilize live and video data, as available, to measure SPOT score ratings by BDOs of behaviors exhibited by passengers against ratings of the same passengers by subject matter experts. This element of the study could help determine whether BDOs are using, or are continuing to use, the SPOT score sheet correctly as time

passes after their initial training. According to S&T officials, the study is to form the basis for BDO performance and training requirements.

The S&T Directorate reported some preliminary findings associated with this research in February 2008. The Directorate reported that although some of the existing literature supported the possibility of using behavioral and physiological cues, the results are not methodologically strong enough to support standardized applications in an operational setting.[39] The preliminary findings also noted that it is not known whether behavioral and physiological cues linked to deception in planning a hostile action will be the same or different as those indicators linked to deception by an individual after they have already engaged in a hostile action. However, an S&T program director stated that although early literature can be characterized as methodologically weak, more recent unpublished research sponsored by DHS, the Department of Defense, and the intelligence community is promising in that it has demonstrated some linkages between behavioral and physiological indicators and deception.[40]

In March 2009, the Under Secretary (Acting) for DHS's S&T Directorate testified that the Directorate had performed an initial validation of the behavior indicators used by BDOs.[41] The Under Secretary stated that this analysis provided statistically significant support that persons demonstrating select behavioral indicators are more likely to possess prohibited items and that behaviors can distinguish deceptive from nondeceptive individuals. According to S&T, this validation was the result of statistical analyses performed by S&T using operational data from the SPOT program database. However, we identified weaknesses in TSA's process for maintaining these data. For example, controls over the SPOT database to help ensure the completeness and accuracy of the data were missing. Specifically, the SPOT database did not have computerized edit checks built into the system to review the format, existence, and reasonableness of data. For example, we found that discrepancies existed

[39]American Institutes for Research, *Behavioral Indicators Related to Deception in Individuals with Hostile Intentions: Interim Results* (Washington, D.C.: February 2008). According to S&T officials, this review included research conducted prior to 2005.

[40]DHS could not provide us with specific contacts related to the sources of this research; we were therefore unable to determine the extent to which it has demonstrated linkages between behavioral and physiological indicators and deception.

[41]Statement of the Under Secretary (Acting), DHS S&T Directorate, before the Subcommittee on Homeland Security, Committee on Appropriations, U.S. House of Representatives, March 26, 2009.

between the number of passengers arrested by local law enforcement at the screening checkpoints and the number of screened passengers recorded as arrested. In another example, we found that the total number of LEO referrals differed from the number of passenger records with information on the reasons for LEO referral. Internal control standards state that controls should be installed at an application's interfaces with other systems to ensure that all inputs are received and are valid and that outputs are correct and properly distributed.[42] TSA officials explained these issues as data anomalies and planned to change instructions to staff entering data to reduce these problems. Although TSA is taking steps to update the SPOT database, which are discussed later in this report, the data used by S&T to conduct its preliminary validation of related behaviors lacked such controls. In addition, BDOs could not input all behaviors observed in the SPOT database because the database limits entry to eight behaviors, six signs of deception, and four types of prohibited items per passenger referred for additional screening. Because of these data-related issues, meaningful analyses could not be conducted to determine if there is an association between certain behaviors and the likelihood that a person displaying certain behaviors would be referred to a LEO or whether any behavior or combination of behaviors could be used to distinguish deceptive from nondeceptive individuals. As a result, TSA lacks assurance that the SPOT data can be used effectively to determine that the person poses a risk to aviation security. S&T has recognized weaknesses in the procedures for collecting data on passengers screened by SPOT and plans to more systematically collect data during its study by, for example, requiring BDOs to record more complete and accurate information related to a passenger referral immediately following resolution.

The S&T study is an important step to determine whether SPOT is more effective at identifying passengers who may be threats to the aviation system than random screening. However, S&T's current research plan is not designed to fully validate whether behavior detection and appearances can be effectively used to reliably identify individuals in an airport terminal environment who pose a risk to the aviation system. For example, research on other issues, such as determining the number of individuals needed to observe a given number of passengers moving at a given rate per day in an airport environment or the duration that such observation can be conducted by BDOs before observation fatigue affects

[42]GAO/AIMD-00-21.3.1.

effectiveness, could provide additional information on the extent to which SPOT can be effectively implemented in airports. In another example, Dr. Ekman told us that additional research could help determine the need for periodic refresher training since no research has yet determined whether behavior detection is easily forgotten or can be potentially degraded with time or lack of use. While S&T officials agree on the need to validate the science of behavior detection programs, they told us that some of these other issues could be examined in the future but are not part of the current plan due to time and budgetary constraints. According to S&T, some additional analysis is underway to inform the current BDO selection process. This analysis is intended to provide information on the knowledge, skills, abilities, and other characteristics of successful BDOs. Since the analysis is scheduled for completion in May 2010, it remains unclear to what extent the findings will help to validate the science related to SPOT. While we recognize the potential benefits of these efforts, we believe that an assessment by an independent panel of experts of the planned methodology of DHS's study could help DHS assess the costs and benefits associated with a more comprehensive methodology designed to fully validate the science related to SPOT. Our prior work has recommended the use of such independent panels for comprehensive, objective reviews of complex issues.[45] In addition, according to the National Research Council, an independent panel could provide an objective assessment of the methodology and findings of DHS's study to better ensure that SPOT is based on validated science. Thus, 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.

[45]See GAO, *Oil and Gas Royalties: The Federal System for Collecting Oil and Gas Revenues Needs Comprehensive Reassessment*, GAO-08-691 (Washington, D.C.: Sept. 3, 2008). GAO, *Combating Nuclear Smuggling: Additional Actions Needed to Ensure Adequate Testing of Next Generation Radiation Detection Equipment*, GAO-07-124T (Washington, D.C.: Sept. 18, 2007). GAO, *Space Operations: NASA Efforts to Develop and Deploy Advanced Spacecraft Computers*, GAO/IMTEC-89-17 (Washington, D.C.: Mar. 31, 1989). GAO, *Quadrennial Defense Review: Future Reviews Could Benefit from Improved Department of Defense Analyses and Changes to Legislative Requirements*, GAO-07-709 (Washington, D.C.: Sept. 14, 2007). GAO, *Coast Guard: Challenges for Addressing Budget Constraints*, GAO/RCED-97-110 (Washington, D.C.: May 1997).

SPOT Was Deployed Nationwide on Basis of Threat, but Without a Comprehensive Risk Assessment

According to DHS's National Infrastructure Protection Plan (NIPP), risk assessments are to be documented, reproducible (so that others can verify the results), defensible (technically sound and free of significant errors), and complete. The NIPP states that comprehensive risk assessments are necessary for determining which assets or systems face the highest risk, for prioritizing risk mitigation efforts and the allocation of resources, and for effectively measuring how security programs reduce risks. For a risk assessment to be considered complete, the NIPP states that it must specifically assess threat, vulnerability, and consequence;[44] after these three components have been assessed, they are to be combined to produce a risk estimate.[45]

According to TSA, SPOT was deployed to TSA-regulated airports on the basis of threat information in TSA's Current Airport Threat Assessment list.[46] TSA deployed SPOT to 161 of 457 TSA-regulated airports. TSA officials told us that this selective deployment creates unpredictability for persons seeking to cause harm to the aviation system because they would not know which airports had BDO teams and because BDOs are occasionally sent out to the smaller airports that do not have BDOs on a permanent basis. Although TSA's selective deployment of SPOT was based on threat information, TSA did not conduct vulnerability and consequence assessments to inform the deployment of BDOs. As a result,

[44]DHS's NIPP defines risk as a function of threat, vulnerability, and consequence. Threat is an indication of the likelihood that a specific type of attack will be initiated against a specific target or class of targets. Vulnerability is the probability that a particular attempted attack will succeed against a particular target or class of targets. Consequence is the effect of a successful attack.

[45]As updated in 2009, the NIPP states that to be complete a risk assessment is to assess threat, vulnerability, and consequence for every defined risk scenario. However, because the original 2006 version of the NIPP described risk assessments that included all three components as "credible," our previous reports use this term rather than "complete" (see GAO, *Transportation Security: Comprehensive Risk Assessments and Stronger Internal Controls Needed to Help Inform TSA Resource Allocation*, GAO-09-492 (Washington, D.C.: Mar. 27, 2009)).

[46]We reported in March 2009 that TSA does not assign uncertainty or varying levels of confidence associated with the intelligence information used to identify threats to the transportation sector and guide TSA's planning and investment decisions. Since TSA's intelligence products have not assigned confidence levels to its analytic judgments, it is difficult for TSA to correctly prioritize its tactics and investments based on uncertain intelligence. In March 2009, we recommended that TSA work with the Director of National Intelligence to determine the best approach for assigning uncertainty or confidence levels to analytic intelligence products and to apply this approach. TSA agreed with this recommendation and said it has begun taking action to address it. See GAO-09-492.

it could not combine the results to conduct a comprehensive risk assessment to inform the deployment of BDOs to those airports with the highest risks.

TSA officials told us that while they have not completed a comprehensive risk assessment for airport security, they have prepared and are currently reviewing a draft of a comprehensive, scenario-based Aviation Modal Risk Assessment—known as the AMRA—which is to serve as a comprehensive risk assessment for aviation security.[47] According to TSA officials, the AMRA is to address all three elements of risk for domestic commercial aviation, general aviation, and air cargo.[48] Although TSA planned to release the AMRA in February 2008, it now expects to finalize the AMRA in 2010. According to TSA, the AMRA may help provide information for the prioritization of BDO deployment within airports, but could not provide specifics on how it would do so. Further, TSA officials noted that information from AMRA would inform BDO deployment in conjunction with other TSA priorities not related to SPOT.[49] Since the AMRA is not yet complete, it is not clear whether it will provide the risk analysis—including assessments of vulnerability and consequence—needed to inform TSA's decisions and planning for any revisions or future deployment of SPOT. If AMRA lacks information relevant to the deployment of SPOT and further research determines that SPOT has a scientifically validated basis for using behavior detection for counterterrorism purposes in the airport environment, then conducting a comprehensive risk assessment of airports could strengthen TSA's ability to establish priorities and make cost-effective resource decisions

[47]The AMRA is being developed by TSA pursuant to the National Strategy for Aviation Security, which was issued according to Homeland Security Presidential Directive-16. HSPD-16 directs the production of an overarching national strategy to optimize and integrate government-wide aviation security efforts. AMRA was previously known as the Air Domain Risk Assessment or ADRA.

[48]Commercial aviation includes that sector of the nation's civil aviation system that provides for the transportation of individuals by scheduled or chartered operations for a fee, including air carriers and airports. General aviation includes all civil aviation other than commercial and military operations, including flight operations such as personal/family transportation, emergency services, wildlife and land surveys, traffic reporting, agricultural aviation, firefighting, and law enforcement. Air cargo is defined as cargo carried on passenger and all-cargo aircraft.

[49]In addition, TSA states that its risk management analysis toolset may also be useful in prioritizing BDO deployments since the toolset analyzes various scenarios for which the use of BDOs may be applicable.

regarding the deployment of BDOs to those airports deemed to have the highest priority risks.

TSA Deployed SPOT Nationwide Without Conducting a Cost-Benefit Analysis but Such an Analysis Could Help Inform Program Decisions Moving Forward

DHS and other federal guidance recommend conducting a cost-benefit analysis before implementing new programs to avoid unnecessary costs and identify the best way to achieve goals at the lowest costs among potential alternatives. Our prior work has also supported the use of cost-benefit analyses during retrospective reviews to validate the agency's original assumptions regarding costs and benefits.[50] In addition, the DHS February 2006 *Cost-Benefit Analysis Guidebook* and OMB guidance both recommend the use of cost-benefit analysis, both in the planning stage for a program, and when significant milestones or financial options are to be assessed.[51] The DHS *Guidebook* states that a cost-benefit analysis is designed to identify optimal financial solutions among competing alternatives. OMB guidance also identifies cost-benefit analysis as one of the key principles to be considered when making capital expenditures, that expected benefits of proposed actions should be explained, and that a baseline should be identified discussing costs and benefits in comparison with clearly defined alternatives. DHS's 2006 and 2009 NIPPs also state that priority is to be given to those protective measures that provide the greatest mitigation of risk for the resources that are available. The DHS NIPPs add that effective protective programs seek to use resources efficiently by focusing on actions that offer the greatest mitigation of risk for any given expenditure. In addition, measuring cost effectiveness of SPOT was a key TSA goal in an October 2005 version of the SPOT strategic plan.

Although the DHS and OMB guidance recommend that a cost-benefit analysis be conducted prior to deploying a program nationwide—and potentially incurring substantial costs—TSA did not conduct such an analysis of SPOT to inform its pilot testing prior to full-scale nationwide deployment. In early 2003, TSA began conducting a pilot test of the SPOT program at Boston Logan airport to better understand the benefits of the program. According to Boston Logan's Federal Security Director, the

[50]See GAO, *Reexamining Regulations: Opportunities Exist to Improve Effectiveness and Transparency of Retrospective Reviews*, GAO-07-791 (Washington, D.C.: July 16, 2007).

[51]DHS, *Cost Benefit Analysis Guidebook* (Washington, D.C.: February 2006); OMB, Circular No. A-94, *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs* (Washington, D.C.: October 1992); OMB, Circular No. A-4, *Regulatory Analysis* (Washington, D.C.: September 2003).

primary purpose of this pilot test was to understand the potential of the program, not to validate its success.[52] TSA officials stated that the program had several benefits, one of which was its "negligible cost." However, TSA did not analyze the pilot test results to determine if SPOT was more cost effective than other alternatives, such as random screening of passengers. In October 2004, TSA implemented additional pilot programs in Providence, Rhode Island and Portland, Maine with the goal of providing Federal Security Directors with an additional layer of security to identify high-risk passengers for additional screening using behavior detection techniques. TSA concluded that the pilot program was successful and cited several security benefits of these pilots. For example, TSA personnel in Providence identified two individuals in possession of illegal drugs, who were then arrested. Law enforcement also arrested another individual referred to them for providing a fraudulent passport. In another example, BDOs in Portland discovered a passenger with multiple passports and a hidden luggage compartment. The passenger was interviewed by LEOs and later released.

TSA determined that these initial pilot tests at three airports were successful without comparing pilot test data to other possible security alternatives. For example, the results of random screening of passengers at the pilot airports could have provided TSA with objective baseline data. Specifically, these data could have been compared to data collected during the SPOT pilots to determine if SPOT was more effective than random screening in detecting passengers who pose a potential risk to aviation security. TSA concluded that the pilot tests were successful because pilot airports were able to easily incorporate SPOT into their security program, train personnel in SPOT, and implement procedures for an additional layer of security according to TSA.

TSA conducted additional pilot tests at the Minneapolis-St. Paul, Minnesota and Bangor, Maine airports in October 2005. TSA also deployed the program to nine additional airports in response to TSA's holiday preparedness plan in December 2005 to further operationally test the program. Senior SPOT program officials explained that TSA did not conduct an analysis of the pilot testing because the program was in its

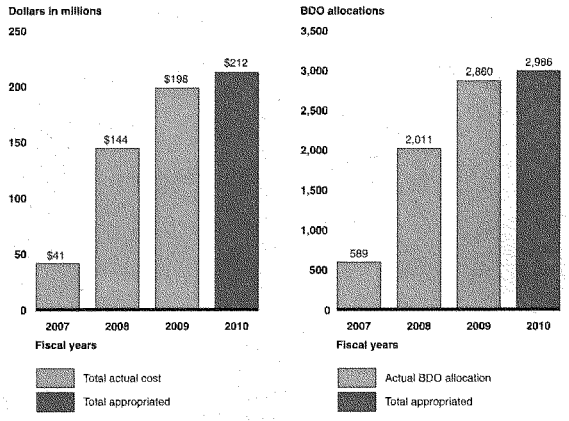
[52]A pilot test is a preliminary test or study to try out procedures and discover problems before the main study begins. This enables researchers to make last minute corrections and adjustments. In a pilot, the entire study with all of its instruments and procedures is conducted in miniature. See W.P. Vogt, *Dictionary of Statistics and Methodology: A Nontechnical Guide for the Social Sciences* (Newbury Park: Sage Publications, 1993).

infancy and officials were focused on deploying SPOT to additional airports. Since that time, TSA has not conducted a cost-benefit analysis, which could help the agency establish the value of the program relative to other layers of aviation security. Moreover, a cost-benefit analysis could also be useful considering recent program growth. For example, from fiscal year 2007 through fiscal year 2009, TSA allotted about $383 million for SPOT. During this period, SPOT's share of TSA's total screening operations budget increased from 1 percent to 5 percent.[53] The conference report accompanying the fiscal year 2010 DHS appropriations act designates $212 million of the appropriated aviation security funding for the SPOT program.[54] A cost-benefit analysis could have provided TSA management with analysis on whether this allocation was a prudent investment, as well as whether this level of investment in SPOT is appropriate. Figure 3 shows the growth in the budget and personnel numbers for SPOT from fiscal years 2007 through 2010.

[53]The increase rate for TSA's other screening operations combined was about 0.27 percent from fiscal year 2007 to fiscal year 2009 (from $3.727 billion to $3.737 billion, a $10 million increase). The screening operations account includes privatized screening; passenger and baggage screener performance, compensation, and benefits; screener training and other; human resource services; and checkpoint support.

[54]See H.R. Rep. No. 111-298, at 77 (2009) (Conf. Rep.).

Figure 3: Budget and Personnel Growth in the SPOT Program, Fiscal Years 2007 through 2010



Source: GAO analysis of TSA data.

Note: The actual BDO allocation for fiscal year 2009 is as of June 2009. The appropriated amount for SPOT for fiscal year 2010 is the amount reflected in the conference report accompanying the fiscal year 2010 DHS appropriations act. The appropriated amounts prior to fiscal year 2010 cannot be determined because funding was appropriated as a lump sum with funding for other screeners and the relevant conference reports did not allocate a specific amount for SPOT. BDO allocation figures are full-time equivalents.

SPOT's Strategic Plan Could be Strengthened by Addressing Key Characteristics of a Successful Strategy

Our previous work,[55] and the Government Performance and Results Act,[56] set forth several key elements of a strategic plan. Such plans can guide agencies in planning and implementing an effective government program. Table 1 summarizes the desirable characteristics of an effective strategic plan, as identified in our prior work. In April 2009, we reported that these characteristics are the starting point for developing a strategic plan.[57]

Table 1: Summary of Desirable Characteristics for Developing a Strategic Plan

Desirable characteristic	Description
Purpose, scope, and methodology	Addresses why the plan was produced, the scope of its coverage, and the process by which it was developed.
Problem definition and risk assessment	Addresses the particular problems and threats the plan is directed towards.
Goals, subordinate objectives, activities, and performance measures	Addresses what the plan is trying to achieve, steps to achieve those results, as well as the priorities, milestones, and performance measures to gauge results.
Resources, investments, and risk management	Addresses what the plan will cost, the sources and types of resources and investments needed, and where resources and investments should be targeted based on balancing risk reductions with cost.
Organizational roles, responsibilities, and coordination	Addresses who will implement the plan, what their roles will be compared to others, and mechanisms for them to coordinate their efforts.
Integration and implementation	Addresses how the plan relates to the agency's other goals, objectives, and activities, to other federal and nonfederal entities involved in implementation or coordination, and their plans to implement the strategic plan.

Source: GAO analysis based on GAO-09-369 and GAO-04-408T.

TSA officials at Boston Logan airport told us that they completed the first strategic plan for SPOT in 2006. The strategic plan was last updated in March 2007. The March 2007 plan includes some of the desirable characteristics described above, such as an overall purpose. However, incorporating additional characteristics of an effective strategic plan could enhance the plan's usefulness in program management and resource allocation decisions to effectively manage the deployment of SPOT if TSA

[55]GAO-04-408T.

[56]Pub. L. No. 103-62, 107 Stat. 285 (1993).

[57]See GAO, *National Preparedness: FEMA Has Made Progress, but Needs to Complete and Integrate Planning, Exercise, and Assessment Efforts*, GAO-09-369 (Washington, D.C.: Apr. 30, 2009).

determines that the program has a scientifically valid basis. TSA officials stated that they believed the plan was sufficiently comprehensive to develop a national program, such as SPOT. However, these officials told us that the plan was not updated after TSA expanded the program in 2008 and 2009. They also stated that the program's focus remained on deploying SPOT to additional airports. Our assessment of the extent to which the SPOT strategic plan addresses these characteristics is presented below.

Purpose, scope, and methodology: The SPOT strategic plan addresses why the plan was developed (i.e., purpose) and the scope of its coverage. Specifically, the plan describes a strategy to utilize behavior detection screening as an additional layer of security. The plan also notes that the primary focus is to expand SPOT in the aviation environment while also developing a capability to deploy BDOs to support security efforts in all modes of transportation. However, the plan does not discuss the process by which it was developed (i.e., methodology). According to TSA, officials responsible for developing the plan received input from relevant stakeholders at Boston Logan airport and TSA headquarters. We believe incorporating the methodology into the plan could make the document more useful to TSA and other organizations, such as local law enforcement, responsible for implementing the plan.

Problem definition and risk assessment: The plan addresses the particular threat it is directed towards. Specifically, the plan describes the need to implement SPOT to counter terrorist activities, improve security, and incorporate additional layers of protection within aviation security. However, the plan does not incorporate risk assessment information to identify priorities or guide program implementation because TSA has not conducted a comprehensive risk assessment related to the deployment of SPOT.[58] Using available risk assessment information to inform the development of a strategic plan would help ensure that clear priorities are established and focused on the areas of greatest need. Specifically, incorporating the results of a risk assessment in the program's strategic plan could help inform TSA's decisions such as whether to deploy SPOT to additional TSA-regulated airports, to shift SPOT teams from one airport to another, or to remove SPOT at airports where the benefit of addressing the risk does not outweigh the costs, as well as to identify and communicate

[58]TSA, *Strategic Plan for Behavior Detection Program* (Washington, D.C.: 2007).

the risks to aviation security if SPOT was not deployed to all TSA-regulated airports.

Goals, subordinate objectives, activities, and performance measures: The plan outlines several goals, objectives, and activities for the SPOT program to achieve. For example, the plan outlines a goal to develop multimodal partnerships, including at the local level, to support SPOT. An associated objective for this goal includes identifying and fostering advocates within each mode of transportation by developing transportation, intelligence, and law enforcement working groups with relevant officials to share information and foster cooperation. The plan also includes a goal to develop and implement performance measures for SPOT. However, the plan did not include performance measures for SPOT. Incorporating performance measures into the plan could help TSA officials measure progress in implementing the plan's goals, objectives, and activities.

Resources, investments, and risk management: The plan does not identify the costs and resources needed to achieve program objectives discussed in the plan. Incorporating information about cost and resources would facilitate TSA's ability to allocate resources across programs according to priorities and constraints, track costs and performance, and shift such investments and resources as appropriate.

Organizational roles, responsibilities, and coordination: The SPOT program relies on a close partnership with law enforcement officers at airports. TSA provides briefings to law enforcement on the SPOT program, and TSA officials conduct outreach efforts to local law enforcement as needed. The SPOT SOP guidance and SPOT training include guidance about ensuring that LEOs receive complete and accurate information about each SPOT referral. However, while the strategic plan identifies TSA officials and offices as responsible parties for implementing the strategic plan, it does not provide guidance on how to effectively link the roles, responsibilities, and capabilities of federal, state, and local officials providing program support. Moreover, although SPOT SOP guidance discusses the need for BDOs to coordinate with other TSA personnel, such as TSOs and TDCs, TSA does not identify their roles and responsibilities in regards to the SPOT program in the program's strategic plan. Integrating these elements into the strategic plan could help to clarify the relationships between these various implementing parties, which would thereby increase accountability and improve the effectiveness of implementation.

Integration and implementation: The SPOT strategic plan does not discuss how its scope complements, expands upon, or overlaps with other related strategic documents. For example, TSA's April 2008 Office of Security Operations Organizational Business Plan for Fiscal Year 2010 describes how its goals—including those for SPOT—relate to DHS and TSA strategic goals.[59] However, TSA does not link goals in the SPOT strategic plan with other related strategic documents, such as the Aviation Implementation Plan of DHS's Transportation Systems Sector-Specific Plan[60] and the Passenger Checkpoint Screening Program Strategic Plan.[61] By linking goals in its SPOT strategic plan to other TSA efforts, TSA could better ensure that the program's objectives are integrated with other TSA security programs and that resources are used effectively by minimizing any unnecessary duplication with these other actions.

More Fully and Consistently Utilizing Available Information Technology Could Enhance TSA's Ability to Identify Threats to the Aviation System

Inconsistencies in the use of available information technology to aid in the collection and recording of data on passengers by BDOs during referrals to LEOs, lack of guidance on, or a mechanism for, BDOs to request the TSA's Transportation Security Operations Center to run the names of passengers exhibiting suspicious behaviors against law enforcement and intelligence databases, and the Center's not checking all of the databases available to it—have limited TSA's ability to identify potential terrorist threats to the aviation system.[62] Among other information, these databases include terrorism-related watch lists.

[59]TSA, Office of Security Operations, *Strategy Deployment Organizational Business Plan for Fiscal Year 2010* (Washington, D.C.: Apr. 8, 2008).

[60]Within the Transportation System Sector-Specific Plan, the aviation implementation plan outlines transportation security goals and key objectives with associated programs within the aviation community. The plan notes that SPOT is intended to identify suspicious activities within the aviation domain.

[61]TSA issued its Passenger Checkpoint Screening Program Strategic Plan in August 2008 to outline its strategy and approach for implementing advanced security capabilities in airport checkpoints using a combination of people, processes, and technology at all airport checkpoints. The plan cites TSA's behavior detection capability as one of three strategic initiatives.

[62]The Transportation Security Operations Center is the central operations and information-gathering point for TSA across the nation; it serves as a 24/7-point of contact for all transportation security concerns, providing access to multiple criminal justice and intelligence-related databases.

Systematic Collection of Data on Passengers Identified Through the SPOT Program Could be Improved to Better Identify Activity Potentially Harmful to the Aviation System

TSA is not fully utilizing the resources it has available to systematically collect the information obtained by BDOs on passengers whose behaviors and appearances resulted in either SPOT referral screening, or in a referral to LEOs, and who thus may pose a risk to the aviation system. TSA's July 2008 Privacy Impact Assessment on the TSA Transportation Security Operations Center, and its August 2008 Privacy Impact Assessment on SPOT, state that information may be obtained by BDOs to check an individual's identity against intelligence, terrorist, and law enforcement databases and to permit intelligence analysts to conduct trend analysis.[63]

The August 2008 SPOT Privacy Impact Assessment states that information about a passenger who has exceeded the SPOT behavior threshold, leading to LEO referral, may be collected and entered into DHS's Transportation Information Sharing System.[64] According to the SPOT Privacy Impact Assessment, information collected may be submitted to the Transportation Information Sharing System database for analysis, and, through it to other linked intelligence databases and the intelligence analysts who study them, to detect, deter, and defeat a criminal or terrorist act in the transportation domain before it occurs. The SPOT Privacy Impact Assessment notes that terrorist acts that threaten transportation security are most vulnerable in the planning stages and that the timely passage of SPOT referral information may assist in identifying such efforts before they become operational. A June 2008 Transportation Information Sharing System Privacy Impact Assessment similarly states that one goal is to use the system data to find trends and patterns that may indicate preoperational terrorist or criminal activity—that is, to "connect the dots" about a planned terrorist attack or criminal enterprise. Information in TSA's Transportation Information Sharing System is primarily activity or behavioral information but may also contain personal information regarding the individuals identified by the BDO through SPOT. According

[63]DHS, *Privacy Impact Assessment for the TSA Operations Center Incident Management System* (Washington, D.C.: July 8, 2008), and *Privacy Impact Assessment for the Screening of Passengers by Observation Techniques (SPOT) Program* (Washington, D.C.: Aug. 5, 2008).

[64]DHS, *Privacy Impact Assessment for the Transportation Information Sharing System* (Washington, D.C.: June 2008). The Transportation Information Sharing System is a database owned by the TSA's FAMS component; the data entered into it may be shared with other federal, state, or local law enforcement and law enforcement support entities. Federal air marshals file reports related to the observation of suspicious activities and input this information, as well as incident reports submitted by airline employees and other individuals within the aviation domain, into the Transportation Information Sharing System.

to TSA, BDOs do not analyze the data obtained during referrals; if they have the appropriate training, they may enter the data by computer into the Transportation Information Sharing System, where they can be analyzed by intelligence analysts. Other appropriately trained and officially designated TSA officials, such as Federal Security Directors, may also enter data into the system.

According to TSA, a 2008 pilot program it conducted that involved BDOs entering data into the Transportation Information Sharing System database was "invaluable," in part because over 40 referrals have since been passed on to other LEO organizations for further investigation, most of which came from BDO input. A February 2006 TSA memorandum describes the Transportation Information Sharing System as "a critical element in the success of SPOT" because it provides the necessary platform for the reporting of information obtained as a result of SPOT referrals. TSA noted that through the use of the Transportation Information Sharing System, two different BDO teams had separately identified and selected the "same extremist" for secondary questioning.[65] TSA officials also told us about an incident in which an individual sought to board an aircraft with a handgun on two separate occasions, at two different airports. Although the handgun was detected both times, the individual was released after providing what seemed to be a credible explanation. After the second incident, however, intelligence analysts who reviewed the system information saw that this individual had tried twice in 2 weeks to bring a weapon onto an aircraft. A LEO was dispatched to the person's home, and an arrest was made. Without the data inputted into the system both times, no pattern would have been detected by the analysts, according to TSA. Although the pilot program illustrated the benefits of BDOs entering data into the system, access to the system was not expanded to all SPOT airports in 2008 or 2009.

Internal control standards call for management to develop policies, procedures, and techniques to help enforce management directives. TSA does not provide official guidance on how or when BDOs or other TSA personnel should enter data into the Transportation Information Sharing System or which data should be entered. Official guidance on what data should be entered into the system on passengers could better position TSA

[65]Because the SPOT program has not been scientifically validated, it cannot be determined if these anecdotal results were better than if passengers had been pulled aside at random, rather than as a consequence of being identified for further screening by BDOs.

personnel to be able to consistently collect information to facilitate synthesis and analysis in "connecting the dots" with regard to persons who may pose a threat to the aviation system.

On March 18, 2010, TSA officials told us that TSA recognizes the value of recording SPOT incidents for the purposes of intelligence gathering. As a result, TSA decided that certain data would be entered into the Transportation Information Sharing System, and would, in turn, be analyzed as a way to potentially "connect the dots" with other transportation security incidents.[66]

TSA officials said that the Federal Security Director at each SPOT airport has been given the discretion to decide which personnel should have access to the Transportation Information Sharing System. However, TSA has not developed a plan detailing how many personnel would have access to the system, or when they would have access at SPOT airports. TSA officials said that training is currently being provided to personnel responsible for using the system at all SPOT airports although they did not provide information on the number being trained.

Standard practices for defining, designing, and executing programs include developing a road map, or program plan, to establish an order for executing specific projects needed to obtain defined programmatic results within a specified time frame. However, TSA stated that it has not developed a schedule or milestones by which database access will be deployed to SPOT airports, or a date by which access at all SPOT airports will be completed. Setting milestones for expanding Transportation Information Sharing System access to all SPOT airports, and setting a date by which the expansion will be completed, could better position TSA to identify threats to the aviation system that may otherwise go undetected and help TSA track its progress in expanding Transportation Information Sharing System access as management intended.

[66]Some details about the process were deleted because TSA considered them to be Sensitive Security Information.

Guidance on and a Mechanism for Running Names of Referred Passengers Through the Databases Available to the Transportation Security Operations Center Could Help Improve SPOT Practices

Internal control standards state that policies, procedures, techniques, and other mechanisms are essential to help ensure that actions are taken to address program risks.[67] The current process makes the BDOs dependent on the LEOs with regard to the timeliness that LEOs respond to BDO calls for service, as well as with regard to whether the LEOs choose to question the passengers referred to them or conduct a background check. Our analysis of the SPOT referral database found a wide variation in the percent of times that LEOs responded to calls for service at SPOT airports.[68] Moreover, if a local LEO decides to run a background check on a passenger referred to them, they would be accessing the FBI's NCIC and not other intelligence and law enforcement databases.

Although LEOs may not always respond to calls for service, question passengers, or check passenger names against databases available to TSA, TSA has not developed a mechanism allowing BDOs to send information to the Transportation Security Operations Center about passengers whose behavior indicates that they may be a possible threat to aviation security. According to TSA's July 2008 Transportation Security Operations Center Privacy Impact Assessment, passenger information may be submitted to the Transportation Security Operations Center to ascertain, as quickly as possible, the individual's identity, whether they are already the subject of a terrorist or criminal investigation, or to analyze suspicious behavior that may signal some form of preoperational surveillance or activity.[69]

Our survey of Federal Security Directors at SPOT airports found a notable inconsistency in the rates at which BDOs at different airports contacted the Transportation Security Operations Center.[70] Developing additional guidance in the SPOT operating procedures could help improve consistency in the extent to which BDOs utilize Transportation Security Operations Center resources. Given the range of responses we received from SPOT airports about whether the BDOs contact the Transportation Security Operations Center to verify passenger identities and run their

[67] GAO/AIMD-00-21.3.1.

[68] Some details from our analysis were deleted because TSA considered them to be Sensitive Security Information.

[69] This information can be submitted about individuals whose suspicious activity resulted in BDO or LEO referral. See TSA's July 2008 Transportation Security Operations Center Privacy Impact Assessment.

[70] Some details of our survey results were deleted because TSA considered them to be Sensitive Security Information.

names against terrorist and intelligence databases and the inconsistencies identified related to LEO responses to BDO requests for service, developing a standard mechanism and providing BDOs with additional guidance could help TSA achieve greater consistency in the SPOT process. Such a mechanism would provide designated TSA officials with a means of verifying passenger identities and help them determine whether a passenger was the subject of a terrorist or criminal investigation and thus posed a risk to aviation security.

Standards for internal control state that effectively using available resources, including key information databases, is one element of functioning internal controls.[71] In this connection, it is widely recognized among intelligence entities and police forces that a capability to "run" names against databases that contain criminal and other records is a potentially powerful tool to both identify those with outstanding warrants and to help discover an ongoing criminal or security-related incident. Additionally, TSA recommended in an April 2008 Organizational Business plan for its Office of Security Operations that the SPOT program should establish a mechanism and policy for allowing real-time checks of federal records for individuals whose behavior indicates they may be a threat to security.[72] The Office of Security Operations plan also states that BDOs should communicate the data to U.S. intelligence centers, with the purpose of permitting rapid communication of this information to local LEOs to take action. However, TSA officials told us that because of safety concerns, the Transportation Security Operations Center does not provide information from database checks directly to BDOs because BDOs are not LEOs, are unarmed, and do not have the training needed to deal with potentially violent persons.[73] If the mechanism discussed in the Office of

[71]See GAO/AIMD-00-21.3.1. For example, information should be recorded and communicated to management and others within the entity who need it and in a form and within a time frame that enables them to carry out their internal control and other responsibilities. Further, effective information technology management is critical to achieving useful, reliable, and continuous recording and communication of information.

[72]TSA, *Strategy Deployment, Organizational Business Plan, Office of Security Operations, Fiscal Year 2010* (Washington, D.C.: April 2008). According to TSA, the Office of Security Operations is the operational arm of TSA and employs the largest TSA workforce. It is responsible for airport checkpoint and baggage screening operations as well as other special programs designed to secure all assigned transportation modes.

[73]In March 2010, TSA told us that over the next 18 months, it will expand access to information classified up to the "Secret" level to an additional 10,000 TSA personnel, including all BDOs, all SPOT Transportation Security Managers (who are responsible for the local operations of the SPOT program and supervision of the BDOs), and all Supervisory TSOs (who directly supervise TSOs and the screening process).

Security Operations business plan were implemented, it would allow the Transportation Security Operations Center to use BDO information to conduct real-time record checks of passengers and communicate the results to LEOs for action. Such a mechanism could increase the chances to detect ongoing criminal or terror plans.

TSA's Transportation Security Operations Center Does Not Use All Database Resources When Contacted

The final report of the National Commission on Terrorist Attacks Upon the United States (the "9/11 Commission Report") recommends that in carrying out its goal of protecting aviation, TSA should utilize the larger set of information maintained by the federal government, that is, the entire Terrorist Screening Database—the U.S. government's consolidated watch list that contains information on known or suspected international and domestic terrorists—as well as other government databases, such as intelligence or law enforcement databases.[74] However, the Transportation Security Operations Center is not using all the resources at its disposal to support BDOs in verifying potential risks to the aviation system. This reduces the opportunities to "connect the dots" that would increase the chances of detecting terrorist attacks in their planning stage, which the SPOT Privacy Impact Assessment states is when they are the most vulnerable.

According to TSA, the Transportation Security Operations Center has access to multiple law enforcement and intelligence databases that can be used to verify the identity of airline passengers; these include among others:[75]

1. the Selectee list, which identifies persons who must undergo enhanced screening at the checkpoint prior to boarding;
2. the No-Fly list,[76] which lists persons prohibited from boarding aircraft; and

[74]The Terrorist Screening Database is the central terrorist watchlist consolidated by the FBI's Terrorist Screening Center and used by multiple agencies to compile their specific watchlists and for screening.

[75]The other databases available to TSA are omitted because TSA considered them to be Sensitive Security Information.

[76]The No-Fly list is used to identify individuals who should be prevented from boarding an aircraft. The No-Fly and Selectee lists contain applicable records from the FBI's Terrorist Screening Center consolidated database of known or suspected terrorists. Pursuant to Homeland Security Presidential Directive 6, dated September 16, 2003, the Terrorist Screening Center—operational since December 2003 under the administration of the FBI—was established to develop and maintain the U.S. government's consolidated terrorist

3. the Terrorist Identity Datamark Environment terrorist list.[77]

TSA stated that the Transportation Security Operations Center checks passenger names submitted to it against these three databases if the passenger has been referred by a BDO to a LEO, but has not been arrested. Of the three databases that the Transportation Security Operations Center is to check in the case of a referral, passengers would have already been screened against two—the Selectee and No-Fly lists—in accordance with TSA passenger prescreening procedures when purchasing a ticket. The third database checked—the Terrorist Identity Datamark Environment—tracks terrorists but not persons wanted for other crimes. The FBI's NCIC information system would contain names of such persons, but is not among the three databases checked for nonarrest referrals. If the passenger has been arrested, the Transportation Security Operations Center will run the passenger's name against the additional law enforcement and intelligence databases available to it.

In addition, TSA told us that the Operations Center does not have direct electronic access to the Terrorist Screening Database and must call the FBI's Terrorist Screening Center to provide it with a name to verify. TSA stated that this is done if a passenger's identity could not be verified using the Operations Center databases. In effect, if a passenger has been referred to a LEO, but not arrested, the Operations Center is to check the three databases shown above to verify the passenger's identity. If a passenger has been arrested, but the three databases do not list the person, the Center can check the additional databases available to it. If none of these databases can verify the person's identity, the Operations Center can contact the Terrorist Screening Center by telephone to request further screening.

screening database (the watch list) and to provide for the use of watch-list records during security-related screening processes. See GAO-08-136T, *Aviation Security: TSA Is Enhancing Its Oversight of Air Carrier Efforts to Screen Passengers against Terrorist Watch-List Records, but Expects Ultimate Solution to Be Implementation of Secure Flight* (Washington, D.C.: Sept. 9, 2008).

[77]According to DHS, the Terrorist Identity Datamark Environment is the database maintained by the National Counterterrorism Center— the primary organization in the U.S. government for integrating and analyzing intelligence pertaining to terrorism and counterterrorism—to serve as a central repository for all information on known or suspected international terrorists with the exception of purely domestic terrorism information. See, DHS, Office of Inspector General, *The DHS Process for Nominating Individuals to the Consolidated Terrorist Watchlist* (Washington, D.C.: February 2008).

For passengers who have risen to the level of a LEO referral at an airport checkpoint, having the Transportation Security Operations Center consistently check their names against all the databases available to it could potentially help TSA identify threats to the aviation system and aid in "connecting the dots." TSA indicated that there are no obstacles to rapidly checking all databases rather than the three listed. We did not analyze the extent to which the law enforcement and intelligence databases available to TSA may contain overlapping information.

TSA Lacks Program Effectiveness Measures for SPOT but Is Taking Steps to Improve Evaluation Capabilities

TSA has established some performance measures by tracking SPOT referral and arrest data, but lacks the measures needed to evaluate the effectiveness of the SPOT program and, as a result, has not been able to fully assess SPOT's contribution to improving aviation security. TSA emphasized the difficulty of developing performance measures for deterrence-based programs, but stated that it is developing additional measures to quantify the effectiveness of the program. The SPOT program uses teams to assess BDO proficiency, provide individual and team guidance, and address issues related to the interaction of BDOs with TSA checkpoint personnel. However, TSA does not systematically track the teams' recommendations or the frequency of the teams' airport visits. TSA states that it is working to address these issues and plans to do so by the end of fiscal year 2010.

TSA Has Taken Action to Collect Data for Some Performance Measures, but Work Remains to Assess Progress Towards Achieving Strategic Goals

TSA agreed that the SPOT program lacked sufficient performance measures in the past, but stated that it has some performance measures in place including tracking data on passengers referred for additional screening and the resolution of this screening, such as if prohibited items were found or if law enforcement arrested the passenger and the reason for the arrest. TSA is also working to improve its evaluation capabilities to better assess the effectiveness of the program. DHS's NIPP, internal controls standards, and our previous work on program assessment state that performance metrics and associated program evaluations are needed to determine if a program works and to identify adjustments that may improve its results.[78] Moreover, standard practices in program management for defining, designing, and executing programs include

[78]DHS, *National Infrastructure Protection Plan: Partnering to Enhance Protection and Resiliency* (Washington, D.C.: 2009); GAO/AIMD-00-21.3.1; and GAO, *Performance Measurement and Evaluation: Definitions and Relationships*, GAO-05-739SP (Washington, D.C.: May 2005).

developing a road map, or program plan, to establish an order for executing specific projects needed to obtain defined programmatic results within a specified time frame.[79] Congress also needs information on whether and in what respects a program is working well or poorly to support its oversight of agencies and their budgets; and agencies' stakeholders need performance information to accurately judge program effectiveness.[80] For example, in the Senate Appropriations Committee report accompanying the fiscal year 2010 DHS appropriations bill,[81] the committee noted that while TSA has dramatically increased the size and scope of SPOT, resources were not tied to specific program goals and objectives. In addition, the conference report accompanying the fiscal year 2010 DHS appropriations act requires TSA to report to Congress, within 60 days of enactment, on the effectiveness of the program in meeting its goals and objectives, among other things.[82] This report was completed on March 15, 2010.

Although TSA tracks data related to SPOT activities including prohibited items, law enforcement arrests related to SPOT referrals, and reasons for the arrests (output measures), it has not yet developed measures to gauge SPOT's effectiveness in meeting TSA strategic goals (outcome measures), such as identifying individuals who may pose a threat to the transportation system.[83] OMB encourages the use of outcome measures because they are more meaningful than output measures, which tend to be more process-oriented or means to an end.[84] For example, TSA's Assistant General

[79] The Project Management Institute, The Standard for Program Management© (2006).

[80] GAO, *Executive Guide: Effectively Implementing the Government Performance and Results Act*, GAO/GGD-96-118 (Washington, D.C.: June 1996).

[81] See S. Rep. No. 111-31, at 56-57 (2009); see also S. Rep. No. 110-396, at 59 (2008).

[82] See H.R. Rep. No. 111-298, at 77 (2009) (Conf. Rep.). The report further directs that GAO review the report submitted by TSA and provide its findings to the committees no later than 120 days after the SPOT report is submitted to the committees.

[83] Output measures help determine the extent to which an activity was performed as planned. Outcome-related measures are more robust measures because they provide a more comprehensive assessment of the success of the agency's efforts, as stated in DHS's 2009 NIPP.

[84] OMB and the Council for Excellence in Government, *Performance Measurement Challenges and Strategies* (Washington, D.C.: June 18, 2003).

Manager for the Office of Operation Process and Performance Metrics[85] told us that SPOT staffing levels are currently used as one performance metric. The official said that since the SPOT is an added layer of security, additional SPOT staffing would add to security effectiveness. While staffing levels may help gauge how fast the program is growing, they do not measure the effectiveness in meeting strategic goals.

Similarly, TSA also cited the number of prohibited items discovered by BDOs in SPOT metrics reports as a measure of program success.[86] However, TSA told us that possession of a prohibited item is often an oversight and not an intentional act; moreover, other checkpoint screening layers are intended to find such items, such as the TSOs and the property screening equipment.[87] TSA also cited measures of BDO job performance as some of the existing measures of program effectiveness, but noted that these are "pass/fail" assessments of individual BDOs, rather than overall program measures.

TSA notes that one purpose of the SPOT program is to deter terrorists, but that proving that it has succeeded at deterring terrorists is difficult because the lack of data has presented challenges for the SPOT program office when developing performance measures. We agree that developing performance measures, especially outcome measures, for programs with a deterrent focus is difficult. Nevertheless, such measures are an important tool to communicate what a program has accomplished and provide information for budget decisions. TSA uses proxy measures—indirect measures or indicators that approximate or represent the direct measure—to address deterrence, other security goals, or a combination of both. For example, TSA tracks the number of prohibited items found and individuals arrested as a result of SPOT referrals. According to OMB, proxy measures are to be correlated to an improved security outcome, and the program should be able to demonstrate—such as through the use of

[85]The Office's primary work involves metrics infrastructure; it assists TSA programs, if requested, in developing applications to track quantitative measures, such as surrendered items. It also tracks data for its Management Objectives Report related to three areas: employees, security effectiveness, and efficiency.

[86]The types of prohibited items found have included knives, guns, gun ammunition, certain chemicals, strike-anywhere matches, and certain liquids/gels/aerosols; other illegal items discovered include narcotics and fraudulent identity documents.

[87]According to TSA, TSOs focus on detecting high-risk threats which have the ability to cause catastrophic damage to an airplane in flight (e.g., explosives).

modeling—how the proxies tie to the eventual outcome.[88] In using a variety of proxy measures, failure in any one of the identified measures could provide an indication on the overall risk to security. However, developing a plan that includes objectives, milestones, and time frames to develop outcome-based performance measures could better position TSA in assessing the effectiveness of the SPOT program.

With regard to more readily quantifiable output performance measures, such as the number of referrals by BDOs, or the ratio of arrests to referrals, TSA was limited in its ability to analyze the data related to these measures. The SPOT database includes information on all passengers referred by BDOs for additional SPOT screening including the behaviors of the passengers that led to the additional screening, as well as the resolution of the screening process (e.g., no further action taken, law enforcement notification, law enforcement investigation, arrested, and reason for arrest). However, TSA reported that any analysis of the data had to be done manually.[89]

In March 2010, TSA migrated the SPOT referral data to its Performance Management Information System, allowing for more statistical and other analyses. According to TSA, migrating the SPOT referral database will enhance the SPOT program's analytic capabilities. For example, TSA stated that it would be able to conduct trend analyses, better segregate data, and create specific reports for certain data. This includes better tracking of performance data at specific airports, analyzing by categories of airports (threat or geographic location), and tracking the performance data of individual BDOs, such as number of referrals, number of arrests, arrest to referral ratios, and other analyses. However, since these changes to the database were not complete at the time of our audit, we could not assess whether the problems we identified with the database had been corrected.

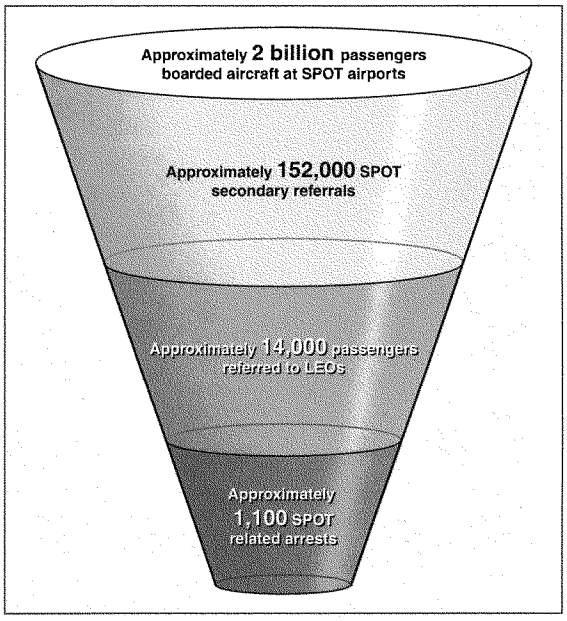
[88]OMB and the Council for Excellence in Government, *Performance Measurement Challenges and Strategies* (Washington, D.C.: June 18, 2003).

[89]We also found that the SPOT referral database had a number of weaknesses. TSA designated our discussion of these weakness as sensitive security information.

Over 4 Years, SPOT Resulted in About 1,100 Arrests Out of Almost 14,000 Referrals to Law Enforcement

The SPOT referral database records the total number of SPOT referrals since May 29, 2004, how many were resolved, how many passengers BDOs referred to LEOs, the recorded reasons for the referral, and how many referrals led to arrests, among other things. As shown in figure 4, we analyzed the SPOT referral data for the period May 29, 2004, to August 31, 2008.

Figure 4: Passenger Boardings at SPOT Airports, May 29, 2004, through August 31, 2008



Source: GAO analysis of TSA and Bureau of Transportation Statistics data.

Note: Figure 4 is not drawn to scale.

Figure 4 shows that approximately 2 billion passengers boarded aircraft at SPOT airports from May 29, 2004, through August 31, 2008.[30] Of these, 151,943 (less than 1/100th of 1 percent) were sent to SPOT referral screening, and of these, 14,104 (9.3 percent) were then referred to LEOs. These LEO referrals resulted in 1,083 arrests, or 7.6 percent of those referred, and less than 1 percent of all SPOT referrals (0.7 percent of 151,943).

We also analyzed the reasons for arrests resulting from SPOT referrals, for the May 29, 2004, through August 31, 2008, period. Table 2 shows, in descending order, the reasons for the arrests.

Table 2: Reasons for Arrests from SPOT Referrals, May 29, 2004 through August 31, 2008

Reasons for arrest	Number	Percentage
Illegal alien	427	39
Outstanding warrants	209	19
Possession of fraudulent documents	166	15
Other	128	12
Possession of suspected drugs	125	12
No reason given	16	1
Undeclared currency	8	1
Suspect documents	4	0
Total	**1,083**	**99***

Source: TSA, SPOT referral database from period of May 29, 2004, through August 31, 2008.

*Total does not add to 100 percent due to rounding.

While SPOT personnel did not determine a specific reason for arrest for 128 cases categorized as "other" or 16 other cases categorized as "no reason given," our analysis of the SPOT database found that a specific reason for arrest could have been determined for these cases by using the LEO resolution notes included in the database. For example, we identified 43 additional arrests related to fraudulent documents, illegal aliens, and suspect documents, among others. The remaining 101 arrests originally characterized as "other" or "no reason given" included arrests for reasons

[30]Our estimate of the total number of passengers who went through checkpoints is based on Bureau of Transportation Statistics data that we obtained for the airports at which SPOT was deployed during this period. Some figures were rounded.

such as intoxication, unruly behavior, theft, domestic violence, and possession of prohibited items. Many of the arrests resulting from BDO referrals would typically fall under the jurisdiction of various local, state, and federal agencies and are not directly related to threats to aviation security. For example, the 427 individuals arrested as illegal aliens, and the 166 arrested for possession of fraudulent documents, are subject to the enforcement responsibilities shared by U.S. Immigration and Customs Enforcement (ICE) and CBP. Although outstanding warrants and the possession of fraudulent or suspect documents could be associated with a terrorist threat, TSA officials did not identify any direct links to terrorism or any threat to the aviation system in any of these cases.

According to TSA, anecdotal examples of BDO actions at airports show the value added by SPOT to securing the aviation system. However, because the SPOT program has not been scientifically validated, it cannot be determined if the anecdotal results cited by TSA were better than if passengers had been pulled aside at random, rather than as a consequence of being identified for further screening by BDOs. Some of the incidents cited by TSA include the following.

- A BDO referred two passengers who were traveling together to referral screening due to suspicious behavior. During secondary screening, one passenger presented fraudulent travel documents. The other could not produce any documentation of his citizenship and it was determined he was in the United States illegally. ICE responded and interviewed both passengers. ICE stated one passenger was also in possession of $10,000 dollars which alarmed positive for narcotics when swept by a K-9 team. ICE arrested one passenger on a federal charge of possession of fraudulent identification documents and entry without inspection. ICE stated charges are still pending for the possession of $10,000. The second passenger was charged with a federal charge of entry without inspection.
- A BDO referred a passenger to referral screening for exhibiting suspicious behavior. Port Authority of Portland (Oregon) Police responded and interviewed the passenger who did not give a statement. LEOs conducted an NCIC check which revealed that there was an outstanding warrant for the failure to appear for a theft charge. LEOs arrested the passenger on a state charge for an outstanding warrant for the failure to appear for theft.
- A BDO referred a passenger for referral screening due to suspicious behavior. During the referral, the passenger admitted that he was unlawfully present in the United States. The Orlando (Florida) Police Department and CBP responded and interviewed the passenger who stated he had $100,000 in his checked baggage, which was confirmed

by CBP. The passenger was arrested on a federal charge of illegal entry.

Because these are anecdotal examples, they cannot be used to reliably generalize about the SPOT program's overall effectiveness or success rate. Our analysis of the SPOT referral database found that the referral data do not indicate if any of the passengers sent to referral screening, or those arrested by LEOs after being referred to them, intended to harm the aircraft, its passengers, or other components of the aviation system. Additionally, SPOT officials told us that it is not known if the SPOT program has ever resulted in the arrest of anyone who is a terrorist, or who was planning to engage in terrorist-related activity.

Reviewing Airport Video Recordings of Individuals Later Arrested or Who Pleaded Guilty for Engaging in Terrorism-Related Activities Could Help TSA Better Identify Terrorist-Related Behaviors

Studying airport video recordings of the behaviors exhibited by persons waiting in line and moving through airport checkpoints and who were later charged with or pleaded guilty to terrorism-related offenses could provide insights about behaviors that may be common among terrorists or could demonstrate that terrorists do not generally display any identifying behaviors. TSA officials agreed that examining video recordings of individuals who were later charged with or pleaded guilty to terrorism-related offenses, as they used the aviation system to travel to overseas locations allegedly to receive terrorist training or to execute attacks, may help inform the SPOT program's identification of behavioral indicators. In addition, such images could help determine if BDOs are looking for the right behaviors or seeing the behaviors they have been trained to observe.

Using CBP and Department of Justice information, we examined the travel of key individuals allegedly involved in six terrorist plots that have been uncovered by law enforcement agencies.[91] We determined that at least 16 of the individuals allegedly involved in these plots moved through 8 different airports where the SPOT program had been implemented.[92] Six

[91]The analysis included only flights leaving the United States. Department of Justice data show that more than 400 individuals have been convicted in the United States for terrorism-related offenses since September 11, 2001. We did not examine the travel itineraries of all these individuals.

[92]The events included the Mumbai, India attack of 2008; a plot to attack the Quantico, Virginia, Marine base in 2008; an effort by five Americans to receive training and fight in Pakistan in December 2009; a plot to attack infrastructure in New York City in 2009; an effort to provide men and support for terrorists in Somalia in 2008; and an attack on a U.S. base in Afghanistan by an American who received training in Pakistan. We were unable to confirm whether BDOs were stationed at the checkpoints used by these individuals at the time they traveled.

of the 8 airports were among the 10 highest risk airports, as rated by TSA in its Current Airport Threat Assessment. In total, these individuals moved through SPOT airports on at least 23 different occasions. For example, according to Department of Justice documents, in December 2007 an individual who later pleaded guilty to providing material support to Somali terrorists boarded a plane at the Minneapolis-Saint Paul International Airport en route to Somalia to join terrorists there and engage in jihad. Similarly, in August 2008 an individual who later pleaded guilty to providing material support to Al-Qaeda boarded a plane at Newark Liberty International Airport en route to Pakistan to receive terrorist training to support his efforts to attack the New York subway system.

Our survey of Federal Security Directors at 161 SPOT airports indicated most checkpoints at SPOT airports have surveillance cameras installed. As we previously reported, best practices for project management call for conducting feasibility studies to assess issues related to technical and economic feasibility, among other things.[83] In addition, Standards for Internal Control state that effectively using available resources is one element of functioning internal controls.[84] TSA may be able to utilize the installed video infrastructure at the nation's airports to study the behavior of persons who were later charged with or pleaded guilty to terrorism-related offenses, and determine whether BDOs saw the behaviors. The Director of Special Operations in TSA's Office of Inspection told us that video recordings could be used as a teaching tool to show the BDOs which behaviors or activities they did or did not observe. In addition, TSA indicated that although the airports may have cameras at the security screening checkpoints, the cameras are not owned by TSA, and in many cases, they are not accessible to TSA. However, TSA officials lack information on the scope of these potential limitations because prior to our work TSA did not have information on the number of checkpoints equipped with video surveillance. We obtained this information as part of our survey of Federal Security Directors at SPOT airports. While TSA officials noted several possible limitations of the use of the existing video surveillance equipment, these images provide TSA a means of acquiring information about terrorist behaviors in the checkpoint environment that

[83]See, GAO, *Supply Chain Security: Feasibility and Cost-Benefit Analysis Would Assist DHS and Congress in Assessing and Implementing the Requirement to Scan 100 Percent of U.S.-Bound Containers*, GAO-10-12 (Washington, D.C.: Oct. 30, 2009). The Project Management Institute, *A Guide to the Project Management Body of Knowledge*.

[84]GAO/AIMD-00-21.3.1.

is not available elsewhere. If current research determines that the SPOT program has a scientifically validated basis for using behavior detection for counterterrorism purposes in the airport environment, then conducting a study to determine the feasibility of using images captured by video cameras could better position TSA in identifying behaviors to observe.

Standardization Teams Assess BDO Proficiency in SPOT Activities and Provide Guidance and Mentoring to BDOs

TSA sends standardization teams to SPOT airports on a periodic basis to conduct activities related to quality control. Teams observe SPOT operations at an airport for several days, working side by side with the BDOs, on multiple shifts, observing their performance, offering guidance, and providing training when required. According to TSA, the purpose of a standardization team visit is to provide operational support to the BDOs, which includes additional training, mentoring, and guidance to help maintain a successful SPOT program.

The standardization teams are comprised of at least two G-Band, or Expert[35] BDOs who have received an additional week of training on SPOT behaviors and mentoring skills. SPOT officials stated that the SPOT program uses its standardization teams to assess overall BDO proficiency by observing BDOs, reviewing SPOT score sheet data, and other relevant data. Standardization teams may also provide a Behavior Observation and Analysis review class to refresh BDOs if the team determines that such a class is needed. The SPOT program director also said that the standardization teams aim to monitor the airport's compliance with the SPOT program's Standard Operating Procedures. As part of this mentoring approach, the standardization teams provide individual and team guidance to the BDOs, offer assistance in program management, and cover issues related to the interaction of BDOs with other TSA checkpoint personnel.

TSA reported to us that it does not systematically track the standardization teams' recommendations or the frequency of the teams' airport visits. Standards for Internal Control state that programs should have controls in place to assess the quality of performance over time and ensure that the findings of audits and other reviews are promptly resolved. Managers are to (1) promptly evaluate findings from audits and other reviews, including those showing deficiencies and recommendations

[35]G-Band, or Expert BDOs, have advanced to a lead role, are able to provide technical expertise on the SPOT program, and are one band away from a supervisory role.

reported by auditors and others who evaluate agencies' operations; (2) determine proper actions in response to findings and recommendations from audits and reviews; and (3) complete, within established time frames, all actions that correct or otherwise resolve the matters brought to management's attention.[86] Although the standardization teams may provide an airport Federal Security Director with recommendations on how to improve SPOT operations, the SPOT program director stated that Federal Security Directors are not required to document whether they have implemented the team recommendations. TSA officials told us that standardization teams can follow up on recommendations made during previous visits. However, TSA did not track whether corrective actions were implemented or the frequency of the team's airport visits to ensure the implementation of the airport's SPOT program. TSA officials stated that they are currently examining ways to compile data to address this issue, and expect to have a system in place in fiscal year 2010.

TSA Developed and Deployed SPOT Training but Further Action Could Enhance Its Effectiveness

Although TSA has taken steps to incorporate all four elements of an effective training program by planning, designing, implementing, and evaluating training for BDOs, further action could help enhance the training's effectiveness. TSA initially consulted outside experts for help in the training's development, which began as a half-day course and has grown to include classroom, on-the-job, and advanced training. TSA also has efforts underway to improve its training program, such as the deployment of SPOT recurrent training. However, TSA evaluations of SPOT program instructors found mixed quality among them, from 2006 onwards. Additionally, TSA has ongoing plans to evaluate the SPOT training for effectiveness, but has not yet developed time frames and milestones for completing the evaluation.

TSA Has Taken Actions to Develop and Deploy SPOT Training

TSA's SPOT Training Evolved Over Time

In 2003, TSA officials at Boston Logan International airport developed the initial half-day training course for SPOT based on an existing course developed for the Massachusetts State Police. Their goal was to take the

[86] GAO/AIMD-00-21.3.1.

behavior detection program designed for law enforcement and apply it to screeners at airport checkpoints. According to TSA officials at Boston Logan, after they recognized that the lecture-style course they originally designed was not effective, they tasked an instructional system designer from TSA's Workplace Performance and Training (the former name of TSA's Operational and Technical Training Division)[97] and an industrial psychologist from the Office of Human Capital to redesign and expand the course, which was piloted in 2005. The 2007 SPOT strategic plan included training objectives for the SPOT program as follows:

- reviewing existing behavior observation training providers,
- establishing and prioritizing multimodal training and assistance efforts based on threat assessments and critical infrastructure,
- establishing a Center of Excellence for Behavior Detection Program training that would continually enhance the quantity and quality of training to selected candidates, and
- developing a recurrent training program designed to refresh and hone skills needed for an effective Behavior Detection Program.

Since that time, the SPOT program implemented, or is in the process of implementing, some of these objectives. For example, in 2008, as part of its effort towards establishing a center for excellence in behavior detection training (third objective), the SPOT program participated in a meeting with behavior detection training officials from various DHS components facilitated by DHS's Screening Coordination Office to promote the sharing of information about behavior detection training and foster future collaboration. Additionally, the SPOT program worked with TSA's Operational and Technical Training Division to create a recurrent training component for BDOs (fourth objective). For example, in 2008, the SPOT program office added a course on detecting microfacial expressions called Additional Behavior Detection Techniques.[98] This 3-day course builds on the behavior detection skills taught in basic training, by teaching

[97]TSA's Operational and Technical Training Division, within the Office of Security Operations, provides assistance with development and implementation of technical training for screening, Behavior Detection Officers, Bomb Appraisal Officers, the Aviation Direct Access Screening Program, and technical management training.

[98]In May 2009, the title of the course was changed to "Additional Behavior Detection Techniques" because ABDT is actually a supplemental tool for BDOs to use during the Casual Conversation phase of SPOT Referral Screening. The course was formerly titled "Advanced Behavior Detection Techniques." Microfacial expressions are very brief facial expressions that can last as little as 1/25 of a second.

BDOs how to detect microfacial expressions. After pilot testing, the course began implementation nationwide in January 2009.

TSA Consulted with Some Experts on Developing SPOT Training

In developing an effective training program, we previously reported that consultation with subject matter experts and expert entities is a core characteristic of the strategic training and development process.[99] TSA SPOT program staff told us that they consulted with experts on behavior detection and observed existing behavior detection courses before deploying the SPOT training program. According to SPOT program officials, a TSA staff member from Boston Logan International Airport attended other training programs offered by other federal agencies and private training organizations to inform the design of SPOT training.[100] TSA officials told us that information from the training courses was used to help develop the list of behaviors or "stress elevators" for the program, and that the point system used to identify passengers for referral screening was based in part on consultations with several subject-matter experts.

TSA documentation also notes that a SPOT working group created in February 2004 consulted with the FBI's Behavioral Science Unit.[101] The Behavioral Science Unit specializes in developing and facilitating training, research, and consultation in the behavioral sciences for the FBI, law enforcement, intelligence, and military communities. While TSA officials from Boston Logan told us that the FBI was included in this initial SPOT working group, these officials agree that coordination with the FBI lapsed until June 2009 when the SPOT Program Office reengaged with the Behavioral Science Unit, and held a meeting with the unit at the FBI Academy in Quantico, Virginia. Since that meeting, a subject matter expert from the SPOT Program Office has been invited to be a member of

[99]GAO, *Human Capital: A Guide for Assessing Strategic Training and Development Efforts in the Federal Government*, GAO-04-546G (Washington, D.C.: Mar. 1, 2004).

[100]The TSA staff member attended the following external training courses: John Reid and Associates' Reid Techniques of Interrogation and Advanced Reid Techniques of Interrogation; Massachusetts State Police Academy's Basic Investigations and Professional Development Program Interview Techniques; International Security Defense Systems' Verification Agent for Virgin Atlantic Security Systems; New Mexico Technology, Materials and Research Center's Prevention and Response to Suicide Bomber Indicators; Abraxis Corporation's Detecting Deception and Eliciting Response; Langevin Learning Services' Instructional Techniques for New Instructors; Ekman Group's Understanding Emotions and Detecting Truth; Chameleon Associates' Suspicious Behavior Detection; and Federal Transit Administration's Terrorist Awareness, Recognition, and Response.

[101]The purpose of the SPOT working group was to help refine the list of SPOT behaviors and to develop standard operating procedures and a concept of operations for the program.

the Terrorism Research and Analysis Project, which is an ongoing working group sponsored by the unit.

In July 2008, DHS's Screening Coordination Office facilitated a collaborative discussion on behavior detection that included TSA, CBP, and Secret Service officials to better ensure that components within DHS share information regarding their efforts in behavior detection and provide a forum for components to have an informed and collaborative discussion on current capabilities, best practices, and lessons learned. According to TSA, no further contact has occurred between the DHS Behavior Detection Working Group and the SPOT program. Thus, the extent to which the working group's expertise will be used to refine or augment SPOT training in the future is not yet clear.

SPOT Program Office Recently Deployed Recurrent Training

Along with basic and remedial training required by the Aviation and Transportation Security Act, TSA policy requires its screening force to regularly complete recurrent (refresher) training. TSA recognized that ongoing training of screeners on a frequent basis and effective supervisory training are critical to maintaining and enhancing skills learned during basic training. According to agency officials, TSA is currently working with DHS S&T to determine the necessary frequency for refresher training for each training course within the SPOT program. Furthermore, TSA plans to place BDOs under TSA's Performance and Accountability Standards System (PASS) beginning in fiscal year 2010. This will include a recertification module.

In 2008, the SPOT program office began the process for developing recurrent SPOT training. Our internal control standards and training assessment guidance suggest that such refresher training should be considered integral to an effective training program from the start because work conditions and environments can be expected to change over time, and additional or updated training is essential to ensuring that the program mission continues to be accomplished.[102] According to the SPOT program office, the recently deployed recurrent training will be semiannual. TSA's Operational and Technical Training Division initially planned to pilot test recurrent training in April 2009 followed by full implementation of the course in approximately May 2009. Because the Operational and Technical Training Division focus was shifted to completing the revisions

[102]GAO/AIMD-00-21.3.1 and GAO-04-546G.

for the SPOT basic certification course, recurrent training was delayed until September 2009 when they released the training on TSA's Online Learning Center.

Instructor Evaluations Found Mixed Quality; Issues with Program Management Led to Instructor Retraining

Our previous work on elements of effective training states that instructors must be both knowledgeable about the subject matter and issues involved, as well as able to effectively transfer these skills and knowledge to others.[103] Moreover, internal control standards state that all personnel need to possess and maintain a level of competence that allows them to accomplish their assigned duties.[104] Management needs to identify appropriate knowledge and skills needed for various jobs and provide needed training, as well as to ensure that those teaching the skills are themselves competent.

TSA conducted internal assessments of SPOT instructors episodically from 2006 through March 2008. These assessments involved a few instructors being rated at a time, and found a wide range of competency among the instructors. In January 2009, TSA's Office of Inspections and Investigations began an investigation of the SPOT training manager, who resigned shortly thereafter. TSA investigators determined that the training manager and other trainers had created a hostile training environment that intimidated some trainees. To address this problem, TSA stated that the program office reexamined the SPOT training program nationally. This included recertifying 47 of 54 SPOT instructors in March 2009, which included evaluation by TSA's Office of Human Capital, Quality Assurance assessors. Additionally, in July 2009, TSA centralized SPOT training at five permanent, regional training facilities in Orlando, Florida; Houston, Texas; Phoenix, Arizona; Denver, Colorado; and Philadelphia, Pennsylvania.[105] According to the SPOT program director, this will allow the SPOT program office more oversight over training. Previously, training was provided at individual airports.

[103]GAO-04-546G.

[104]GAO/AIMD-00-21.3.1.

[105]The SPOT program retains the discretion to train BDOs at a site other than one of the five training facilities if it is more fiscally responsible to do so. For example, if there are 15 BDO candidates at a single airport, the SPOT program will train them at that airport rather than sending them to a training facility.

After the March 2009 recertification training, ratings scores of SPOT instructors showed less variation than did previous ratings. We reviewed the quality assurance instructor evaluations of two SPOT instructors conducted by TSA's Office of Human Capital, Training Standards and Evaluation Branch, and the 167 SPOT program instructor evaluations of 54 SPOT instructors conducted by the SPOT program office and TSA's Operational and Technical Training Division since the program started in October 2006.[106] After the recertification training, 93 percent of instructors were rated as exceeding expectations, compared to 30 percent in the 2006 to September 2008 ratings. Table 3 shows the ratings of instructors for March 2009 compared to the period of 2006 to September 2008.[107]

Table 3: SPOT Instructor Evaluation Ratings, 2006 to September 2008, and March 2009

	Number of instructor evaluations	Unsatisfactory (0-74%)		Needs improvement (75-84%)		Meets expectations (85-94%)		Exceeds expectations (95-100%)		No numeric score given	
		Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
2006 - Sept 2008	73	3	4%	5	7%	36	49%	22	30%	7	10%
March 2009	94	0	0%	1	1%	6	6%	87	93%	0	0%

Source: GAO analysis of TSA Quality Assurance Instructor Evaluations for SPOT.

In addition to the variation in numeric scores and rating levels for the 2006 to September 2008 period, as shown in table 3, we found substantial variation in the comments about instructor competency for the same

[106]Some SPOT instructors have been evaluated multiple times. While the SPOT program office provided us with print or electronic copies of all SPOT instructor evaluations, some forms contained only numeric ratings and no written comments; others had no numeric scores. Because instructor names were redacted from the evaluations, the numbers may include duplicates. Additionally, the evaluations containing written comments were not always filled out using complete sentences, making it difficult to ascertain the rater's assessment of the instructor.

[107]SPOT Instructors are evaluated using a Quality Assurance Instructor Evaluation, TSA Form 1909. Using this form, the evaluator assigns either 0 (zero) points, 0.5 points, or 1 point for each of 57 ratable items depending on whether the instructor meets the standard as written, needs improvement to meet the standard, or does not meet the standard. The total points are then entered into a formula that generates a percentage. This percentage is used to determine the overall rating. Instructors receiving a score of 95 percent to 100 percent are rated as exceeds expectations; 85 percent to 94 percent are rated as meets expectations; 75 percent to 84 percent are rated as needs improvement; and 0 percent to 74 percent are rated as unsatisfactory.

period. For example, in 32 out of 74 instructor evaluation forms that we reviewed where comments were made about the instructor prior to 2009, the comments ranged from superb to needs more experience as an instructor, as well as needs more time performing the job as a BDO to be able to teach others. In the comments on an instructor who was rated as "meets expectations," the instructor was described as having "limited experience within the SPOT program," that this was "a major concern," and it was recommended that the instructor spend as much time as possible functioning as a BDO. In other cases, however, SPOT instructors were described as competent, solid, and outstanding. For example, one instructor who received a rating of "exceeds expectations" was described as a superb instructor who "is a valued member of the National Training Team." As noted above, following the March 2009 recertification training, 93 percent of the instructors received a rating of "exceeds expectations" with only 1 percent "needing improvement." Of the 94 instructor evaluations completed in March 2009, 82 contained written comments. Of these, multiple SPOT instructors were described as excellent, knowledgeable, and effective. For example, an instructor who received a rating of "exceeds expectations" was noted as demonstrating a high degree of material knowledge and great presentation skills. TSA attributed the increase in instructor ratings to two factors. The first is low turnover among SPOT instructors, which allows instructors to hone both their technical and instructor skills. The second factor cited by TSA is that TSA conducted a 2-day instructor refresher training immediately prior to the evaluations in March 2009. To ensure all instructors were reevaluated within a specific time frame, evaluations were scheduled and conducted in a controlled environment. Instructors knew in advance they were going to be evaluated and delivered modules of the BDO certification course to other BDO instructors.

TSA Has Taken Some Action, but Has Not Evaluated the SPOT Training Program for Effectiveness

We previously reported that evaluation is an integral part of training and development efforts, and that agencies need to systematically plan for and evaluate the effectiveness of training and development.[108] Employing systematic monitoring and feedback processes can help by catching potential problems at an early stage, thereby saving valuable time and resources that a major redesign of training would likely entail. Similarly, in 2006, TSA's Operational and Technical Training Division issued general evaluation standards for training programs, stating that training programs

[108]GAO-04-546G.

should be comprehensively evaluated on a periodic basis to identify program strengths and weaknesses.[109] Moreover, standard practices in program management for defining, designing, and executing programs include developing a road map, or program plan, to establish an order for executing specific projects needed to obtain defined programmatic results within a specified time frame.[110]

The former SPOT training manager told us that the SPOT program internally evaluates the effectiveness of SPOT training through the job knowledge tests that BDO candidates must pass following the classroom portion of the training and the SPOT Proficiency/On-the-Job Training Checklist following the on-the-job portion of the training. Furthermore, the former training manager told us that TSA knows that the SPOT training is effective because BDOs are able to recognize behaviors at the checkpoint, and because of BDOs' demonstrated ability to identify criminals—such as drug couriers or people with outstanding arrest warrants—through the screening process.

Although TSA has not conducted a comprehensive analysis of the effectiveness of the SPOT training program, TSA's Office of Human Capital, Training Standards and Evaluation Branch conducted training evaluations to assess how students use what they were taught in the SPOT basic training course. Specifically, from July through September 2008, the Training Standards and Evaluation Branch conducted evaluations at 5 of the 161 airports where the SPOT program is currently operating. Based on BDO feedback at the 5 airports, the Training Standards and Evaluation Branch's final report contained a series of recommendations for improving the SPOT training program. These recommendations and TSA's actions to address them are summarized in table 4.

[109]TSA, Operational and Technical Training Division, Training Standards (Sept. 28, 2006).

[110]The Project Management Institute, The Standard for Program Management© (2006).

Table 4: TSA Training Standards and Evaluation Branch Recommendations for Improving SPOT Training and TSA Actions on the Recommendations

Training Standards and Evaluation Branch recommendations	TSA action on recommendations
Ensure training instructors adhere to a set of professional guidelines.	TSA sent 47 TSA Approved Instructors for the SPOT program to recertification training in March 2009.
Add local policies and procedure as an addendum to the (SPOT) Training.	No action.[a]
Include more role-playing and scenarios in the classroom training so all trainees can practice casual conversation skills.	TSA added more role-playing scenarios to their basic SPOT training.
Develop recurrent training that can be placed on the TSA Online Learning Center.	TSA developed and deployed recurrent training on the TSA Online Learning Center in September 2009.
Develop templates for writing reports.	TSA added an Incident Report Writing course to the TSA Online Learning Center. Additionally, TSA has developed templates for Incident Reports and After Action Reports. TSA has also developed Online Learning Center training for completing SPOT Referral Reports.
Provide more reel world videos.	TSA revised the SPOT training videos in late 2008.
Provide recurrent training of behaviors through online videos.	The video scenarios for recurrent training will be available in the second quarter of fiscal year 2010.
Add parts of the Bomb Appraisal Officer task into the training.	No action.[a]
Provide recurrent training outside of TSA (more Immigration and Customs Enforcement, DEA, and CBP training).	No action.[a]
Have BDOs spend more time with an On-the-Job-Training mentor.	No action.[a]
Validate the training for course content and On-the-Job-Training.	In 2009, in coordination with DHS S&T, TSA began the scientific analysis of the BDO position to empirically derive and validate the knowledge, skills, and attributes that it requires. The analysis is projected to be completed in fiscal year 2010.
Clarify SPOT's "Walk-the-Line" policy and communicate it to all BDO personnel.	TSA issued revised SPOT Standard Operating Procedures to all BDOs in January 2009.

Source: TSA, Training Standards and Evaluation Branch, Office of Human Capital, Memorandum For Operational and Technical Training, and Behavior Detection and Document Validation Branch, Office of Security Operations on Training Transfer (L3) of SPOT Training, October 30, 2008.

*According to TSA, the SPOT program office will determine if the recommended action is appropriate after the BDO job task analysis and training task analysis are completed.

Additionally, in conjunction with S&T, TSA conducted a training effectiveness evaluation on the Additional Behavior Detection Techniques course, which showed a statistically significant increase in knowledge and skills following completion of the course.

S&T is currently conducting a BDO job task analysis, which may be used to evaluate and update the SPOT training curriculum. Following the completion of the job task analysis—anticipated in mid-May 2010—TSA's Operational and Technical Training Division intends to conduct an in-depth training gap analysis,[111] which will take approximately 2 months to complete. Following completion of the training gap analysis, the agency will develop project plans, including milestones for future development efforts, to address any training concerns. However, to date, the agency does not have an evaluation plan including time frames and milestones for completion. According to the Operational and Technical Training Division, TSA will conduct periodic evaluations as the BDO position evolves. By conducting a comprehensive evaluation of the effectiveness of its training program, TSA will be in a better position to determine if BDOs are being taught the knowledge and skills they need to perform their job. Furthermore, by developing milestones and time frames for conducting such evaluations systematically, as well as on a periodic basis, TSA could help ensure that the SPOT training program is evaluated in accordance with its directives to help ensure that the program continues to provide BDOs with the necessary tools required to carry out their responsibilities.

Conclusions

TSA developed the SPOT program in the wake of September 11, 2001, in an effort to respond quickly to potential threats to aviation security by identifying individuals who may pose a threat to aviation security, including terrorists planning or executing an attack who were not likely to be identified by TSA's other screening security measures. Because TSA did not ensure that SPOT's underlying methodology and work methods were scientifically validated prior to its nationwide deployment, an independent panel of experts could help determine whether a scientific foundation exists for the way in which the SPOT program uses behavior detection analysis for counterterrorism purposes in the aviation environment.

[111]The training gap analysis identifies gaps in the training curriculum.

With approximately $5.2 billion devoted to screening passengers and their property in fiscal year 2009, it is important that TSA provides effective stewardship of taxpayer funds ensuring a return on investment for each layer of its security system. As one layer of aviation security, the SPOT program has an estimated projected cost of about $1.2 billion over the next 5 years if the administration's requested funding of $232 million for fiscal year 2011 remains at this level.[112] The nation's constrained fiscal environment makes it imperative that careful choices be made regarding which investments to pursue and which to discontinue. If an independent expert panel determines that DHS's study is sufficiently comprehensive to determine whether the SPOT program is based on valid scientific principles that can be effectively applied in an airport environment for counterterrorism purposes, then conducting a comprehensive risk assessment including threat, vulnerability, and consequence could strengthen TSA's ability in making resource allocation decisions and prioritizing its risk mitigation efforts. Moreover, conducting a cost-benefit analysis could help TSA determine whether SPOT provides benefits greater than or equal to other security alternatives and whether its level of investment in the SPOT program is appropriate. Revising its strategic plan for SPOT to incorporate risk assessment information, cost and resource analysis, and other essential components could enhance the plan's usefulness to TSA in making program management and resource allocation decisions to effectively manage the deployment of SPOT.

Providing guidance on how to use TSA's resources for running passenger names against intelligence and criminal databases available to the Transportation Security Operations Center and helping DHS to connect disparate pieces of information using the Transportation Information Sharing System and other related intelligence and crime database and data sources could better inform DHS and TSA regarding the identity and background of certain individuals and thereby enhance aviation security. In addition, implementing the steps called for in the TSA Office of Strategic Operations plan to provide BDOs with a real-time mechanism to verify passenger identities and backgrounds via TSA's Transportation Security Operations Center could strengthen their ability to rapidly verify the identity and background of passengers who have caused concern, and increase the likelihood of detecting and disrupting potential terrorists

[112]This estimate assumes that there would be no further increases for SPOT over the next 5 years above the requested $232 million level for fiscal year 2011. However, to stay even with inflation, the allocation would likely increase somewhat each year.

intending to cause harm to the aviation system. Additionally, developing outcome-oriented performance measures, making improvements to the SPOT database, and studying the feasibility of utilizing video recordings of individuals as they transited checkpoints and who were later charged with or pleaded guilty to terrorism-related offenses, could help TSA evaluate the SPOT program, identify potential vulnerabilities, and assess the effectiveness of its BDOs. Further, developing a plan for systematic and periodic evaluation of the training provided to BDOs along with time frames and milestones for its completion could help ensure that the SPOT training program is evaluated in accordance with its directives to help ensure that the program continues to provide BDOs with the necessary tools required to carry out their responsibilities.

Recommendations for Executive Action

To help ensure that SPOT is based on valid scientific principles that can be effectively applied in an airport environment, we recommend that the Secretary of Homeland Security convene an independent panel of experts to review the methodology of the DHS S&T Directorate study on the SPOT program to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program. This assessment should include appropriate input from other federal agencies with expertise in behavior detection and relevant subject matter experts.

If this research determines that the SPOT program has a scientifically validated basis for using behavior detection for counterterrorism purposes in the airport environment, then we recommend that the TSA Administrator take the following four actions:

- Conduct a comprehensive risk assessment to include threat, vulnerability, and consequence of airports nationwide to determine the effective deployment of SPOT if TSA's ongoing Aviation Modal Risk Assessment lacks this information.
- Perform a cost-benefit analysis of the SPOT program, including a comparison of the SPOT program with other security screening programs, such as random screening, or already existing security measures.
- Revise and implement the SPOT strategic plan by incorporating risk assessment information, identifying cost and resources, linking it to other related TSA strategic documents, describing how SPOT is integrated and implemented with TSA's other layers of aviation security, and providing guidance on how to effectively link the roles, responsibilities, and capabilities of federal, state, and local officials providing program support.

- Study the feasibility of using airport checkpoint-surveillance video recordings of individuals transiting checkpoints who were later charged with or pleaded guilty to terrorism-related offenses to enhance understanding of terrorist behaviors in the airport checkpoint environment.

Concurrent with the DHS S&T Directorate study of SPOT, and an independent panel assessment of the soundness of the methodology of the S&T study, we recommend that the TSA Administrator take the following six actions to ensure the program's effective implementation:

- To provide additional assurance that TSA utilizes available resources to support the goals of deterring, detecting, and preventing security threats to the aviation system, TSA should:
  - Provide guidance in the SPOT Standard Operating Procedures or other TSA directive to BDOs, or other TSA personnel, on inputting data into the Transportation Information Sharing System and set milestones and a time frame for deploying Transportation Information Sharing System access to SPOT airports so that TSA and intelligence community entities have information from all SPOT LEO referrals readily available to assist in "connecting the dots" and identifying potential terror plots.
  - Implement the steps called for in the TSA Office of Security Operations Business plan to develop a standardized process for allowing BDOs or other designated airport officials to send information to TSA's Transportation Security Operations Center about passengers whose behavior indicates that they may pose a threat to security, and provide guidance on how designated TSA officials are to receive information back from the Transportation Security Operations Center.
  - Direct the TSA Transportation Security Operations Center to utilize all of the law enforcement and intelligence databases available to it when running passenger names, for passengers who have risen to the level of a LEO referral.

- To better measure the effectiveness of the program and evaluate the performance of BDOs, TSA should:
  - Establish a plan that includes objectives, milestones, and time frames to develop outcome-oriented performance measures to help refine the current methods used by Behavior Detection Officers for identifying individuals who may pose a risk to the aviation system.
  - Establish controls to help ensure completeness, accuracy, authorization, and validity of data collected during SPOT screening.

- To help ensure that TSA provides BDOs with the knowledge and skills needed to perform their duties, TSA should:
  - Establish time frames and milestones for its plan to systematically conduct evaluations of the SPOT training program on a periodic basis.

Agency Comments and Our Evaluation

We provided a draft of our report to DHS and TSA on March 19, 2010, for review and comment. On May 3, 2010, DHS provided written comments, which are reprinted in appendix II. In commenting on our report, DHS stated that it concurred with 10 of our recommendations and identified actions taken, planned, or under way to implement them. However, the actions DHS reported it plans to take and has underway do not fully address the intent of our first recommendation. DHS also concurred in principle with an eleventh recommendation stating that it had convened a working group to determine the feasibility of implementing it. DHS commented on the scientific basis underlying SPOT and on two statements in our report that it believed were inaccurate—specifically, DHS disagreed with our reliance on a 2008 National Research Council report published under the auspices of the National Academy of Sciences on issues related to behavior detection, and second, on issues related to unpublished research they had cited as a partial validation of some aspects of the SPOT program.[113] Finally, DHS commented on our conclusion regarding the use of the SPOT referral data.

Regarding our first recommendation that DHS convene an independent panel of experts to review the methodology of DHS's Science and Technology Directorate (S&T) study on SPOT, and to include appropriate input from other federal agencies with relevant expertise, DHS concurred and stated the current process includes an independent review of the program that will include input from other federal agencies and relevant experts. Although DHS has contracted with the American Institutes for Research to conduct its study, it remains unclear who will oversee this review and whether they are sufficiently independent from the current research process. DHS's response also does not describe how the review currently planned is designed to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program. As we noted in our report, research on other issues, such as determining

[113] The National Research Council is a component of the National Academy of Sciences, a part of a private, nonprofit institution, the National Academies, which provide science, technology, and health policy advice under a congressional charter.

the number of individuals needed to observe a given number of passengers moving at a given rate per day in an airport environment or the duration that such observation can be conducted by BDOs before observation fatigue affects effectiveness, could provide additional information on the extent to which SPOT can be effectively implemented in airports. Dr. Paul Ekman, a leading research scientist in the field of behavior detection, told us that additional research could help determine the need for periodic refresher training since no research has yet determined whether behavior detection is easily forgotten or can be potentially degraded with time or lack of use. Thus, questions exist as to whether behavior detection principles can be reliably and effectively used for counterterrorism purposes in airport settings to identify individuals who may pose a risk to the aviation system. To help ensure an objective assessment of the study's methodology and findings, DHS could benefit from convening an independent panel of experts from outside DHS to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program.

DHS also concurred with our second recommendation to conduct a comprehensive risk assessment to determine the effective deployment of SPOT. DHS stated that TSA's Aviation Modal Risk Assessment is designed to evaluate overall transportation security risk, not deployment strategies. However, DHS noted that TSA is in the process of conducting an initial risk analysis using its risk management analysis tool and plans to update this analysis in the future. However, it is not clear from DHS's comments how this analysis will incorporate an assessment of TSA's deployment strategy for SPOT.

DHS also concurred with our third recommendation to perform a cost-benefit analysis of SPOT. DHS noted that TSA is developing an initial cost-benefit analysis and that the flexibility of behavior detection officers already suggests that behavior detection is cost-effective. However, it is not clear from DHS's comments whether its cost-benefit analysis will include a comparison of the SPOT program with other security screening programs, such as random screening, or already existing security measures as we recommended. Completing its cost-benefit analysis and comparing it to other screening programs should help establish whether the SPOT program is cost-effective compared to other layers of security.

With regard to our fourth recommendation to revise and implement the SPOT strategic plan using risk assessment information, DHS concurred and noted that analysis facilitated by the risk management analysis tool

will allow the program to revise the SPOT strategic plan to incorporate the elements identified in our recommendation.

DHS also concurred with our fifth recommendation to study the feasibility of using airport checkpoint-surveillance video recordings to enhance its understanding of terrorist behaviors. DHS noted that TSA agrees this could be a useful tool and is working with DHS's S&T Directorate to utilize video case studies of terrorists, if possible. These cases studies could help TSA determine what behaviors had been demonstrated by these persons convicted of terrorist-related offenses who went through SPOT airports, and what could be learned from the observed behaviors.

DHS concurred with our sixth recommendation that TSA provide guidance in the SPOT SOP or other directives to BDOs, or other TSA personnel, on how to input data into the Transportation Information Sharing System database. DHS stated that the SPOT SOP is undergoing revision, and that the revised version will provide guidance directing the input of BDO data into the Transportation Information Sharing System. DHS anticipates release of the updated SPOT SOP in fiscal year 2010. DHS also agreed that TSA should set milestones and a time frame for deploying Transportation Information Sharing System access to SPOT airports so that TSA and intelligence community entities have information from all SPOT LEO referrals readily available to assist in "connecting the dots" and identifying potential terror plots. DHS stated that TSA is currently drafting a plan to include milestones and a time frame for deploying System access to all SPOT airports.

DHS concurred with our seventh recommendation to develop a standardized process to allow BDOs or other designated airport officials to send information to TSA's Transportation Security Operations Center about passengers whose behavior indicates they may pose a threat to security, and to provide guidance on how designated TSA officials are to receive information back from the Center. DHS stated that TSA has convened a working group to address this recommendation. Moreover, TSA is developing a system and procedure for sending and receiving information from the Center and stated that it anticipates having a system in place later in fiscal year 2010.

DHS concurred in principle with regard to our eighth recommendation that the Transportation Security Operations Center utilize all of the databases available to it when conducting checks on passengers who rise to the level of a LEO referral against intelligence and criminal databases. DHS stated that TSA has convened a working group to address this

recommendation. According to DHS, this group will conduct a study during fiscal year 2010 to determine the feasibility of fully implementing this recommendation. As such, the study is to review the various authorities, permissions, and limitations of each of the databases or systems cited in our report. DHS stated that access to some of the systems, requires more justification than a BDO referral. Further, according to DHS, because some of the databases or systems contain classified information, TSA will also need to adopt a communication strategy to transmit the passenger information between the BDO and Transportation Security Operations Center. DHS stated that TSA will work on a process to collect the passenger information, verify the passenger's identity, through checks of databases, and analyze that information to determine if the passenger is the subject of an investigation and may pose a risk to aviation security.

With regard to our ninth recommendation to establish a plan with objectives, milestones, and time frames to develop outcome-oriented performance measures for BDOs, DHS concurred and stated that TSA intends to consult with experts to develop outcome-oriented performance measures.

DHS also concurred with our tenth recommendation to establish controls for SPOT data. DHS noted that TSA established additional controls as part of the SPOT database migration to TSA's Performance Management Information System and is exploring an additional technology solution to reduce possible errors. As noted in our report, since these changes to the database were not complete at the time of our audit, we could not assess whether the problems we identified with the database had been corrected.

Regarding our eleventh recommendation to establish time frames and milestones to systematically evaluate the SPOT training program on a periodic basis, DHS concurred and stated that TSA intends to develop such a plan following completion of DHS's S&T Directorate's BDO Job Task Analysis, and TSA's training gap analysis, which identifies gaps in the training curriculum.

DHS also commented on the scientific basis underlying SPOT. Specifically, DHS stated that decades of scientific research has shown the SPOT behaviors to be "universal in their manifestation." However, according to DHS, its S&T Directorate is examining the extent to which behavior indicators are appropriate for screening purposes and lead to appropriate and correct security decisions. DHS also commented that the results of this work, which is currently underway, will establish a scientific

basis of the extent to which the SPOT program instruments and methods are valid. Thus, DHS's comments suggest that additional research is needed to determine whether these behaviors can be used in an airport environment for screening passengers to identify threats to the aviation system.

Moreover, DHS took issue with our use of a report from the National Research Council of the National Academy of Sciences stating that we improperly relied upon this report.[114] We disagree. DHS questioned the findings of the National Research Council report and stated that it lacked sufficient information for its conclusions because it principally focused on privacy as it relates to data mining and behavioral surveillance and was not intended to represent an exhaustive or definitive review of the research or operational literature on behavioral screening, including recent unpublished DHS, defense, and intelligence community studies. DHS also stated that the National Research Council report did not study the SPOT program and that the researchers did not conduct interviews with SPOT personnel.

As we noted in our report, although the National Research Council report addresses broader issues related to privacy and data mining, a senior Council official—and one of the authors of the study—stated that the committee included behavior detection as a focus because any behavior detection program could have privacy implications. This official added that the primary objective of the report was to develop a framework for sound decision making for programs, such as SPOT, and help ensure a sound scientific and legal basis. According to this official, the National Academy of Sciences' Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals—which had oversight of the report—was briefed on the SPOT program as part of

[114]National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008). The report's preparation was overseen by the National Academy of Science's 21-member Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals. We reviewed the approach used and the information provided in this study and found the study to be credible for our purposes. The contributors included recognized experts across a variety of fields, including William J. Perry, former Secretary of Defense, and Dr. Tara O'Toole, then-CEO and Director of the Center for Biosecurity of the University of Pittsburgh Medical Center, Professor of Medicine and of Public Health at the University of Pittsburgh. (Dr. O'Toole was subsequently nominated and confirmed as the Under Secretary of the DHS Science and Technology Directorate.)

the study. The Committee also conducted meetings with three experts in behavior detection as part of their research. During the course of our review, we interviewed three Committee members responsible for developing the report's findings, as well as four other behavior detection experts, including the three who participated in the National Research Council study. Our discussions with these experts corroborated the report's findings. Thus, we believe that our use of the Council report was an appropriate and a necessary part of our review.

However, the National Research Council report was only one of many sources that we analyzed with regard to the science of behavioral and physiological screening, and its applicability to an airport environment. As we noted in the description of our methodology, our study included interviews with officials from DHS as well as several of its components and other U.S. government agencies—each of which use elements of behavior detection in their daily work. We also interviewed El Al airline officials, a former director of security at Israel's Ben-Gurion airport, and seven nationally recognized experts in behavior detection as part of our review. Moreover, as we explained in the discussion of our scope and methodology, we conducted a survey about the SPOT program of all 118 Federal Security Directors for all SPOT airports, and conducted site visits to 15 SPOT airports. In addition, we analyzed the SPOT referral database, to the extent the data permitted, covering a 4-year period and the results from 2 billion passengers passing through SPOT airports. Moreover, we attended both the basic and advanced training courses in behavior detection provided by TSA to BDOs, in order to better understand how the program is carried out. Therefore, our analysis of the program was not derived from or based on a single study by the National Research Council as DHS suggested, but rather is based on all of the information we gathered and synthesized from multiple, diverse, expert sources, each of which provided different perspectives about the program, as well as about behavior detection in general.

DHS also disagreed with the accuracy of a statement included in our report that noted DHS S&T could not provide us with specific contacts related to sources of information for certain research it cited as support for the SPOT program. In its comments, DHS stated that it had provided us with all requested documents that represent DHS's S&T Directorate-sponsored research. We agree. However, DHS did not provide us with contact information for the sources of unpublished studies by the Department of Defense and other intelligence community studies that DHS S&T had cited as support for the SPOT program. Without such

information, we are unable to verify the contents of these unpublished studies.

Finally, DHS stated that while we were unable to use the SPOT referral data to assess whether any behavior or combination of SPOT behaviors could be used to reliably predict the final outcome of an incident involving the use of SPOT, it was able to analyze the SPOT referral database successfully after working with TSA to verify scores assigned to different indicators. Our concern with the data did not involve the question of whether some behaviors were entered erroneously, nor whether errors in coding were excessive or non-random. Rather, we were concerned with whether the data on behaviors were complete. Specifically, it cannot be determined from the SPOT referral database whether all behaviors observed were included for each referred passenger by each BDO or whether only the behaviors that were sufficient for a LEO referral were recorded into the database. It is not possible to determine from the database if the number of observed behaviors entered for a given passenger was the total number of observed behaviors, or whether additional behaviors were observed. A rigorous analysis of the relative effects of the different behaviors on the outcomes of the use of SPOT would require each BDO to record, for each of the observable behaviors, whether it was or was not observed.

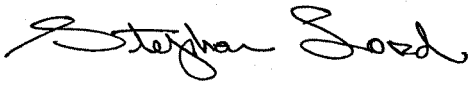
TSA also provided technical comments that we incorporated as appropriate.

We will send copies of this report to the Secretary of Homeland Security; the TSA Administrator (Acting); and interested congressional committees as appropriate. The report will also be available at no charge on the GAO Web site at http://www.gao.gov.

If you or your staff have any questions about this report, please contact me at (202) 512-4379 or lords@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page

of this report. Key contributors to this report are acknowledged in appendix III.

Sincerely yours,



Stephen M. Lord
Director, Homeland Security and Justice Issues

Appendix I: Scope and Methodology

To determine the extent to which the Transportation Security Administration (TSA) determined whether the Screening of Passengers By Observation Techniques (SPOT) program had a scientifically-validated basis for identifying passengers before deploying it, we reviewed literature on behavior analysis by subject matter experts, interviewed seven experts in behavior analysis, interviewed other federal agencies and entities about how they use behavior detection techniques, and analyzed relevant reports and books on the topic. These included a 2008 study by the National Research Council of the National Academy of Sciences that has a discussion regarding deception and behavioral surveillance, as well as other issues related to behavioral analysis.[1] We interviewed Dr. Herbert S. Lin, who was a primary author of the report, as well as Dr. Robert W. Levenson, and Dr. Stephen E. Fienberg, both members of the Academy committee that oversaw the report, about the report's findings with regard to behavior detection, and the extent to which behavior detection in a complex environment, such as an airport terminal, has been validated with regard to its effectiveness in identifying persons who may be a risk to aviation security. Other behavior detection experts we consulted were Dr. Paul Ekman;[2] Dr. Mark Frank;[3] Dr. David Givens;[4] Dr. David Matsumoto;[5]

[1]National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008). The report's preparation was overseen by the NAS's 21-member Committee on Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals. We reviewed the approach used and the information provided in this study and found the study to be credible for our purposes. The contributors included recognized experts across a variety of fields, including William J. Perry, former Secretary of Defense, and Dr. Tara O'Toole, then-CEO and Director of the Center for Biosecurity of the University of Pittsburgh Medical Center, Professor of Medicine and of Public Health at the University of Pittsburgh. (Dr. O'Toole was subsequently nominated and confirmed as the Under Secretary of DHS's Science and Technology Directorate. The National Research Council is a component of the National Academy of Sciences, a part of a private, nonprofit institution, the National Academies, which provide science, technology, and health policy advice under a congressional charter

[2]Dr. Ekman is professor emeritus of psychology at the University of California Medical School, San Francisco, and is considered one of the world's foremost experts on facial expressions. His books include: *Emotions Revealed: Recognizing Faces and Feelings to Improve Communications and Emotional Life* (New York: Holt and Company, 2003); *Emotion in the Human Face* (New York: Pergamon Press, 1972); *Unmasking the Face: A guide to Recognizing Emotions from Facial Clues* (Englewood Cliffs, N.J.: Prentice-Hall, 1975). Dr. Ekman has published more than 100 articles.

[3]Dr. Frank is Associate Professor, Department of Communication, College of Arts and Sciences, at the University at Buffalo, State University of New York. He is on the Advisory Board of the University's Center for Unified Biometrics and Sensors, and has conducted research supported by DHS, the Defense Advanced Research Projects Agency, and the National Science Foundation.

and Mr. Rafi Ron, former director of security at Israel's Ben-Gurion Airport. Dr. Ekman, Dr. Frank, and Mr. Ron provided expert advice for the National Research Council study. Dr. Givens was identified by TSA as having been their principal source for the nonverbal behavior indicators used by the SPOT program. We also interviewed Dr. Lawrence M. Wein, an expert in emergency responses to terror attacks and mathematical models in operations management.[6] In addition, we interviewed officials from the Department of Homeland Security's (DHS) Science and Technology (S&T) Directorate regarding their ongoing research into behavior detection. Although the views of these experts cannot be generalized across all experts in behavior analysis because we selected individuals based on their publications on behavioral analysis or related topics, their recognized accomplishments and expertise, and, in some cases, TSA's use of their work or expertise to design and review the SPOT program's behaviors, they provided us with an overall understanding of the fundamentals of behavior analysis, and how it could be applied.

To determine the basis for TSA's strategy to develop and deploy SPOT and evaluate to what extent SPOT was informed by a cost-benefit analysis and a strategic plan, we reviewed program documentation, including briefings prepared by the SPOT program office during the course of developing and fielding SPOT, two versions of a strategic plan for SPOT, and the 2009 SPOT standard operating procedures guidance. We compared the plans and analyses used by TSA to develop and implement SPOT to criteria on how to develop and implement programs in DHS's 2006 *Cost Benefit Analysis Guidebook*,[7] as well as to Office of Management and Budget guidance on the utility of cost-benefit analyses in program

[4]Dr. Givens is the director of the nonprofit Center for Nonverbal Studies, in Spokane, Washington. He is the author of *Love Signals: A Practical Field Guide to the Body Language of Courtship* (St. Martin's, New York, 2005) and *Crime Signals: How to Spot a Criminal Before You Become a Victim* (St. Martin's, 2008). The Center's Web site links to Dr. Givens' reference tool, *The Nonverbal Dictionary of Gestures, Signs and Body Language Cues*.

[5]Dr. Matsumoto is a Professor, Department of Psychology at San Francisco State University, and is an associate of Dr. Ekman.

[6]Dr. Wein is the Paul E. Holden Professor of Management Science at the Graduate School of Business, Stanford University. His homeland security-related work includes four papers in *Proceedings of the National Academy of Sciences*, on an emergency response to a smallpox attack, an emergency response to an anthrax attack, a biometric analysis of the US-VISIT Program, and an analysis of a bioterror attack on the milk supply.

[7]DHS, *Cost Benefit Analysis Guidebook* (Washington, D.C.: February 2006).

implementation.[8] We also analyzed the development of SPOT in light of the standards and criteria cited in DHS's 2006 National Infrastructure Protection Plan. We met with relevant TSA officials to discuss these issues. To assess whether DHS developed an effective strategic plan for SPOT prior to implementing the program, we interviewed TSA officials involved in development of the SPOT strategic plan. We analyzed whether the SPOT plan incorporated the desirable characteristics of an effective strategic plan as identified by previous GAO work on what strategic plans should include to be considered effective, such as a risk assessment, cost and resources analysis, and a means for collaboration with other key entities.[9] We also examined it in light of the requirements of the Government Performance and Results Act of 1993, which specifies the elements of strategic plans for government programs.[10] We assessed whether the SPOT strategic plan was followed by TSA. As part of our analysis of the planning for SPOT before it was implemented on a nationwide basis, we reviewed TSA documentation related to the development and pilot testing of SPOT, such as a TSA white paper on SPOT, and interviewed key program officials from both headquarters and field offices.[11]

We also interviewed cognizant officials from other U.S. government agencies and agency entities that utilize behavior detection in their work, including U.S. Customs and Border Protection (CBP), the U.S. Secret Service, the TSA's Federal Air Marshal Service (FAMS) component, and the Federal Bureau of Investigation (FBI). We sought their views on the utility of various behavior detection methods, their experience with practicing behavior detection, and asked them about the extent to which TSA had consulted with them in developing and implementing the SPOT program.

To better understand how SPOT incorporated expertise about the use of behavior detection in an airport setting, we interviewed officials from

[8]Office of Management and Budget (OMB), Circular No. A-94, *Guidelines and Discount Rates for Benefit-Cost Analysis of Federal Programs* (Washington, D.C.: October 1992); and Circular A-4, *Regulatory Analysis* (Washington, D.C.: September 2003).

[9]GAO-04-408T.

[10]Pub. L. No. 103-62, 107 Stat. 285 (1993).

[11]TSA, *Screening of Passengers by Observation Technique (SPOT) White Paper for the Department of Homeland Security* (Washington, D.C.: Feb. 22, 2005).

Israel's El Al Airlines, which is cited by TSA as having provided part of the basis of the SPOT program. We asked about El Al's methods to ensure the security of its passenger aircraft, and also interviewed a former head of security at Israel's Ben-Gurion airport, who has advised TSA on security issues. We asked TSA and SPOT program officials about their consultations with El Al, and about the ways in which they had utilized El Al's expertise, as well as about any other entities whose expertise they may have adopted into SPOT.

To determine the challenges, if any, that emerged during implementation of the SPOT program, we interviewed headquarters and field personnel about how the program has utilized the resources available to it to ensure that it is effective. These resources included the support of law enforcement officers (LEOs), to whom passengers are referred by Behavior Detection Officers (BDOs) for additional questioning. In addition, we interviewed SPOT program and TSA officials about the databases available to them at TSA's Transportation Security Operations Center to determine if a suspect passenger is being sought by other U.S. law enforcement or intelligence entities, and whether there is guidance for BDOs on when and how to contact the Transportation Security Operations Center. We also asked about whether there is guidance and training for BDOs on how to access TSA's Transportation Information Sharing System database, which is owned by FAMS, and is available through the Transportation Security Operations Center.[12] To determine if any management challenges had emerged related to management controls in developing and implementing SPOT, we compared TSA's approach for implementing and managing the SPOT program with GAO's *Standards for Internal Control in the Federal Government*[13] and with risk management principles we had previously identified.[14] Our legal counsel office reviewed court decisions relevant to the SPOT program. In addition, we interviewed attorneys from the American Civil Liberties Union, and obtained and reviewed TSA's Privacy Impact Assessments for SPOT, the Transportation Security Operations Center, and the Transportation

[12]The data from interviews of suspicious passengers by FAMS are inputted into the Transportation Information Sharing System, as are reports sent to FAMS from airline employees about suspicious passengers.

[13]GAO/AIMD-00-21.3.1.

[14]GAO, *Transportation Security: Comprehensive Risk Assessments and Stronger Internal Controls Needed to Help Inform TSA Resource Allocation*, GAO-09-492 (Washington, D.C.: Mar. 27, 2009).

Information Sharing System. We also met with and discussed relevant privacy and legal issues with TSA's Offices of Privacy and Civil Rights/Civil Liberties. To obtain data about certain aspects of the SPOT program that the SPOT program office did not have, we conducted a survey of Federal Security Directors[15] whose responsibilities included security at all 161 SPOT airports at the time of our survey. (Some Federal Security Directors have responsibility for more than one airport.) We obtained a 100 percent response rate. This survey asked, among other things, about whether there were cameras at security checkpoints that record the interactions of Transportation Security Officers (TSO), BDOs, and passengers; if the airport authority had an agreement with TSA that specifies certain law enforcement actions during a SPOT referral; and if there was an agreement, or any other comparable guidance that specified a time limit for LEOs to come to checkpoints after being called for help by BDOs.

To determine the extent to which TSA has measured SPOT's effect on aviation security, we obtained and analyzed the TSA SPOT referral database, which records all incidents in which BDOs refer passengers to secondary, more intensive questioning, and which also records all incidents in which BDOs chose to refer passengers to LEOs. We found that the SPOT database was sufficiently reliable to count the number of arrests resulting from referrals from BDOs to LEOs, for examining the reasons for each arrest, and for counting the percentage of times that LEOs responded to BDO calls for service, and the length of time required. Use of these data required us to resolve apparent contradictions and anomalies in the database to make the data useable. Because of data problems, we were unable to conduct analyses to assess whether any behavior or combination of behaviors could be used to predict the final outcome of an incident involving the use of SPOT. In addition, we reviewed relevant standardization team reports and observed a standardization team visit in operation.

In addition, we spoke with BDO managers, Federal Security Directors, and Assistant Federal Security Directors to determine how BDOs are evaluated. To do so, we conducted site visits to 15 commercial airports at which BDOs and SPOT have been deployed, or almost 10 percent of the

[15]Federal Security Directors are the highest ranking TSA security officials at U.S. airports; Assistant Federal Security Directors are their assistants. Both are responsible for all aspects of security at airports, including coordination with federal and nonfederal law enforcement entities operating at airports, such as FAMS, the Drug Enforcement Administration, and CBP.

161 airports with SPOT. We chose these airports taking into account the following criteria, among others: (1) each airport had BDOs deployed, and at each, the SPOT program had been in effect for no less than 3 months; (2) airports were chosen to provide a variety of sizes, as measured in annual passenger volume; physical location within the country (northeast, southwest, central, Pacific Coast, rural, urban); and estimated risk of terrorist incident, using DHS's Current Airports Threat Assessment[16] list (visiting 6 that were in the top 10, and others much lower); (3) BDOs who are employed by contractors, rather than employed directly by TSA; and (4) airports with LEOs who were identified to us by TSA as having received some form of behavior detection training and airports where they were not known to have received such training. In addition, we took into account the location of the airports with regard their proximity to subject matter experts on behavior detection whom we wished to interview, as well as the time and cost required to reach certain airports.

At each of the airports we visited, we interviewed cognizant officials, including the Federal Security Director or Assistant assigned to the airport, the BDO program manager, one or two BDOs, and one or two LEOs who have interacted with BDOs. Since each of these airports differs in terms of passenger volume, physical size and layout, geographic location, and potential value as a target for terrorism, among other things, the results from these visits are not generalizable to other airports. However, these visits provided helpful insight into the operation of SPOT at airports.

In addition, to determine if individuals had transited SPOT airports who were later charged with or pleaded guilty to terrorism-related offenses, we reviewed information contained in (1) the Treasury Enforcement Communication System II database maintained by CBP;[17] (2) Department of Justice information and court documents, including indictments and related documents; and (3) media accounts of individuals accused of

[16]The Current Airports Threat Assessment is a threat estimate designed to provide a snapshot of the current terrorist threat to airports in the United States as well as for major international airports serving as last points of departure for U.S. airlines.

[17]The Treasury Enforcement Communication System was designed to provide controlled access to a large database of information about suspects and to interface with a number of other law enforcement systems. These capabilities are provided to users through various applications, including the Inspection/Interagency Border Inspection System applications that facilitate passenger processing through the implementation of innovative border control technology.

terrorism-related activities. We compared information pertaining to these individuals' dates of transit to the dates when SPOT was deployed to the various airports identified in the Treasury Enforcement Communication System and Justice Department data to determine if SPOT had been deployed at a given airport when the transits occurred. Further, we used our survey of Federal Security Directors at SPOT airports to determine the extent to which video surveillance cameras are present at checkpoints.

To assess the extent that SPOT training incorporates the attributes of an effective training program, we had training experts at TSA headquarters complete a training assessment tool that we developed using our prior work for assessing training courses and curricula.[18] To address training-related issues, including to understand better how other entities train their employees in behavior detection, and what their curricula include, we conducted site visits to the Secret Service, FAMS, CBP, and the FBI, and also interviewed nongovernmental experts on behavior detection (our selection of these experts is discussed above). As part of our assessment of SPOT training, we attended the basic SPOT training course given to BDOs, as well as the advanced SPOT course on behavior detection. We interviewed BDOs and BDO managers about the SPOT training, as well as officials of El Al airlines, with regard to how El Al trains and tests its personnel who utilize behavior recognition and analysis as part of their assessment of El Al passengers.

We conducted this performance audit from May 2008 through May 2010, in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient, appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained provides a reasonable basis for our findings and conclusions based on our audit objectives.

[18]GAO, *Human Capital: A Guide for Assessing Strategic Training and Development Efforts in the Federal Government*, GAO-04-546G (Washington, D.C.: Mar. 1, 2004).

Appendix II: DHS Comments

U.S. Department of Homeland Security
Washington, DC 20528

May 3, 2010

Mr. Steve Lord
Director, Homeland Security & Justice Issues
U.S. Government Accountability Office (GAO)
441 G Street, NW
Washington, DC 20548

Dear Mr. Lord:

Thank you for the opportunity to review and comment on GAO-10-157SU, the draft report titled: *Aviation Security: Efforts to Validate Aspects of TSA's Screening of Passengers by Observation Techniques (SPOT) Program Underway, But Opportunities Exist to Strengthen Validation and Address Operational Changes*. The Transportation Security Administration (TSA) appreciates the U.S. Government Accountability Office's work in planning and conducting its review and issuing this report.

TSA deployed the SPOT program in an effort to mitigate the threat of individuals with potentially hostile intent from boarding a commercial airplane and causing harm. Congress has encouraged the use of behavior recognition to enhance aviation security and has provided resources to support its implementation and expansion. The SPOT program fulfills the mandate of Section 1611 of the Implementing Recommendations of the 9/11 Commission Act, P.L. 110-53, that "TSA shall provide advanced training to the transportation security officers for the development of specialized security skills, including behavior observation and analysis ... in order to enhance the effectiveness of layered transportation security measures."

Intelligence continues to show there is no specific terrorist profile. In a March 10, 2010, hearing before the Senate Homeland Security and Governmental Affairs Committee, TSA Acting Administrator Gale Rossides highlighted the challenge faced by TSA leaders in "balancing the requirement to screen all passengers and to actually focus our officers' attention on the right passengers." TSA designed SPOT to increase its ability to focus on the "right passengers" by identifying persons exhibiting behaviors and appearances that may indicate stress, fear, and deception, and distinguish them from other travelers.

SPOT is Based on Scientific Research and Law Enforcement Practices

TSA's development and deployment of SPOT was a planned and deliberate process based on more than 3 years of operational test-bed assessment of SPOT at Boston's Logan International Airport from June 2003 until nationwide rollout began in fiscal year (FY) 2007. TSA carefully developed SPOT by using selective behaviors recognized within both the scientific and law

www.dhs.gov

enforcement communities as displaying stress, fear, and deception. A SPOT working group, made up of various TSA and U.S. Department of Homeland Security (DHS) components,[1] was created in February 2004. Other organizations, such as the Massachusetts State Police, the Federal Bureau of Investigation (FBI) Behavioral Sciences Unit, and the Federal Law Enforcement Training Center, were also involved in SPOT development. Through these working groups, TSA has developed and finalized SPOT standard operating procedures (SOPs) for a common ability to assess behaviors indicating hostile intent for both aviation and mass transit modes of transportation. TSA continues to consult with its SPOT working group partners as it updates the procedures and science behind the program.

Decades of scientific research have shown the behaviors to be universal in their manifestation. In fact, the DHS Science and Technology Directorate (S&T) completed a study on suicide bomber indicators in July 2009 that illustrates a very high degree of overlap between operationally reported suicide bomber indicators and TSA SPOT behaviors. This result further bolsters TSA's contention that the SPOT program draws from the best practices of many defense, intelligence, and law enforcement organizations.

SPOT Scientific Validation is Ongoing

S&T began research in 2007 to examine the validity of the SPOT program. The series of studies involved in this research is designed to assess the validity of the SPOT scoring system, including the use of individual behavioral indicators to identify high-risk travelers. More specifically, S&T's research plan aims to examine the extent to which these behavioral indicators are appropriate for screening purposes and lead to appropriate and correct security decisions. When this study is complete, SPOT will be one of the most, if not the most, rigorously tested behavior-based security screening programs in existence.

Results of this work will establish a scientific basis of the extent to which the SPOT program, including its instrument and methods, such as the SPOT Referral Report and SOPs, are valid. Although it is challenging to establish the validity of a deterrent program in which the outcomes of interest are extremely rare, critical elements of reliability and validity will be rigorously assessed. Of particular importance is the evaluation of criterion-related validity, or the extent to which travelers are correctly selected for screening based on the SPOT scoring system. Establishing this degree of classification accuracy justifies the use of the SPOT program to discriminate high-risk travelers from low-risk travelers. Regardless of any other metrics, the extent to which the SPOT scores accurately identify high-risk travelers is critically important to program validity.

Following criterion-related validity, the next central element of validity is the consistency of implementation of the instrument and program. This will be examined in a variety of ways, including an investigation of the consistency in the operational use of SPOT behavioral indicators Behavior Detection Officers (BDOs) and across locations and time periods, all of which represents reliability assessment. Finally, construct-related validity, or the extent to which

[1] Includes TSA's Office of Civil Rights, Office of Chief Counsel, and Privacy Office, and DHS's Policy Office and Transportation Security Laboratory.

the SPOT program behaviors truly represent the expressions of high-risk travelers, will be examined by comparing the SPOT behaviors to similar instruments in use for the same purpose. S&T's July 2009 study of suicide bomber indicators was the first step in evaluating construct-related validity.

This research is expected to be completed in FY 2011. TSA understands that after this validation is complete, there will be other areas where further research should be conducted, and it is TSA's intention to complete this research.

National Academy of Sciences (NAS) Report Does Not Represent an Exhaustive or Definitive Review of the Research or Operational Literature on Behavioral Screening

TSA would like to specifically address a few comments in the GAO-10-157SU report that we believe are inaccurate. The report draws heavily from a National Academy of Sciences (NAS) report which is being improperly relied upon. As the sponsor of the NAS study, DHS S&T questioned its findings, stating that the study lacked sufficient information for its conclusions because the NAS study principally focused on privacy as it relates to behavioral surveillance–not on behavioral surveillance technology itself. The study was not intended to, and the results do not represent an exhaustive or definitive review of the research or operational literature on behavioral and physiological screening, including recent findings from unpublished DHS, defense, and intelligence community studies. Furthermore, it should be noted that the report did not study the SPOT program, nor did any of the researchers conduct interviews with SPOT program personnel.

Additionally, GAO states that "DHS S&T could not provide us with specific contacts related to the sources of this research." This statement is not accurate. The record should reflect that DHS S&T provided all requested documents that represented S&T-sponsored research and for which S&T possessed the requisite release authority. DHS was not able to release specific documents related to research for which it was not the originator.

The report further states that the audit team was unable to use the SPOT referral data to assess whether any behavior or combination of SPOT behaviors could be used to reliably predict the final outcome of an incident involving the use of SPOT. However, DHS S&T was able to successfully conduct some preliminary analysis of the SPOT referral database. Prior to analysis of the SPOT reports, S&T worked with TSA to verify the scores assigned to each indicator with the SPOT score sheets and to rescore the pertinent sections and total accordingly for nearly 100,000 operational reports from 2008. While random errors were noted, errors in large databases that require manual entry are not uncommon. Convention suggests that large databases like this typically include an error rate of 3 to 5 percent. As long as such errors are random, the analytical method is robust enough to account for random errors in this range.

In conclusion, TSA strongly believes that behavior detection is a vital layer in its aviation security strategy, and will continue to strengthen as the program matures. Leaders within the community of behavior detection researchers agree. TSA appreciates GAO's work to identify opportunities to enhance the SPOT program, and we will continue to work diligently to address

the issues identified by GAO. Our ongoing progress demonstrates our commitment to TSA's mission of securing our Nation's transportation systems.

We also appreciate the opportunity to provide you with, in collaboration with DHS S&T, comments to GAO's audit recommendations.

Recommendation 1: To help ensure that SPOT is based on valid scientific principles that can be effectively applied in an airport environment, we (GAO) recommend that the Secretary of Homeland Security convene an independent panel of experts to review the methodology of the S&T Directorate study on the SPOT program before the study is implemented to determine whether the study's methodology is sufficiently comprehensive to validate the SPOT program. This assessment should include appropriate input from other federal agencies with expertise in behavior detection and relevant subject matter experts.

Concur. The U.S. Department of Homeland Security (DHS) Science & Technology Directorate's (S&T) current validation process includes an independent and comprehensive review of the ongoing SPOT study to be conducted in support of and in collaboration with the TSA SPOT program. The assessment will include input from other Federal agencies with expertise in behavior detection and relevant subject matter experts. S&T will work with TSA to present the SPOT validation project to the panel, produce a report summarizing the panel's recommendations, and implement pertinent suggestions in FY 2010.

GAO further recommends that if this research determines that the SPOT program has a scientifically validated basis for using behavior detection for counterterrorism purposes in the airport environment, then the TSA Administrator take the following actions:

Recommendation 2: Conduct a comprehensive risk assessment to include threat, vulnerability, and consequence of airports nationwide to determine the effective deployment of SPOT if TSA's ongoing Aviation Modal Risk Assessment lacks this information.

Concur. TSA's Aviation Modal Risk Assessment (AMRA) is designed to evaluate the transportation security risk landscape and compare it to other modes. However the AMRA does not evaluate risk effectiveness of countermeasures or optimal deployment strategies. For the Aviation mode, TSA uses the Risk Management Analysis Tool (RMAT), a risk simulation model based on laboratory and operational data that evaluates risk using threat inputs, vulnerability information, and consequence estimates. TSA is in the process of conducting an initial risk analysis on the SPOT program using RMAT. The risk analysis is based on the initial SPOT validation results and will be updated as the validation study results are finalized.

Recommendation 3: Perform a cost-benefit analysis of the SPOT program including a comparison of the SPOT program with other security screening programs, such as random screening, or already existing security measures.

Concur. The SPOT program will use RMAT to perform a cost-benefit analysis of Behavior Detection Officers (BDOs) as a countermeasure. The first step in the process is the initial risk

5

assessment that is being conducted on the SPOT program using RMAT. For the cost-benefit analysis, costs will be defined as the 5-year total cost of the countermeasure across the aviation system. Benefit will be defined as risk-reduction across the aviation security system against a portfolio of scenarios. TSA is currently developing an initial cost-benefit analysis for a variety of passenger-screening countermeasures including BDOs using the RMAT tool as a basis for analysis. BDOs' flexibility across a variety of risk scenarios suggests that behavior detection is a cost-effective countermeasure.

Recommendation 4. Revise and implement the SPOT strategic plan by incorporating risk assessment information, identifying cost and resources, linking it to other related TSA strategic documents, describing how SPOT is integrated and implemented with TSA's other layers of aviation security, and providing guidance on how to effectively link the roles, responsibilities and capabilities of federal, state, and local officials providing program support.

Concur. The RMAT risk analysis of the BDO program is assisting the SPOT program in identifying other countermeasure capabilities that are linked to the behavior detection capability. This analysis will allow the SPOT program to develop a revision to the SPOT strategic plan that will incorporate the elements identified in the recommendation.

Recommendation 5: Study the feasibility of using airport checkpoint-surveillance video recordings of individuals transiting checkpoints, and who were later charged with or pleaded guilty to terrorism-related offenses, to enhance its understanding of terrorist behaviors in the airport checkpoint environment.

Concur. TSA will study the feasibility of using checkpoint surveillance video recordings of individuals transiting checkpoints, and who were later charged with or pleaded guilty to terrorism-related offenses. TSA agrees that this could be a useful tool in understanding terrorist behaviors in the checkpoint environment.

Additionally, TSA is currently working with DHS S&T/Human Factors to conduct operational video validation of the SPOT program. TSA will use a variety of video case studies to validate the SPOT program including, if possible, reviewing video of terrorists transiting the TSA checkpoint. It is exceedingly rare, however, for video cameras to capture terrorists transiting TSA checkpoints. Unfortunately, this factor significantly reduces the feasibility of conducting these case studies.

GAO also recommends that concurrent with the DHS S&T Directorate study of SPOT, and an independent panel assessment of the soundness of the methodology of the S&T study, the TSA Administrator take the following actions:

Recommendation 6: Provide guidance in the SPOT SOP or other TSA directive to BDOs, or other TSA personnel, on inputting data into the Transportation Information Sharing System (TISS) and set milestones and a timeframe for deploying Transportation Information Sharing System access to SPOT airports so that TSA and intelligence community entities have information from all SPOT Law Enforcement officer (LEO)

referrals readily available to assist in "connecting the dots" and identifying potential terror plots.

Concur. TSA is currently undergoing a revision of the SPOT Standard Operating Procedure (SOP). The SOP will provide guidance directing the input of BDO data into TISS. TSA anticipates release of the updated SPOT SOP in FY 2010. Additionally, TSA is currently drafting a formal plan to include milestones and a timeframe for deploying TISS access to all SPOT airports.

Recommendation 7: Implement the steps called for in the TSA Office of Security Operations Business plan to develop a standardized process for allowing BDOs or other designated airport officials to send information to TSA's Transportation Security Operations Center (TSOC) about passengers whose behavior indicates that they may pose a threat to security, and provide guidance on how designated TSA officials are to receive information back from the Transportation Security Operations Center.

Concur. TSA has convened a working group made up of members of the Office of Security Operations, Office of Chief Counsel, Office of Intelligence, and the Office of Law Enforcement/Federal Air Marshal Service (FAMS) to address this recommendation. TSA is developing a system and procedure for sending and receiving information from the TSOC and anticipates having a system in place in FY 2010. It should be noted that information from BDO referrals has been transmitted to the TSOC previously; however, TSA agrees to institute a standardized process.

Recommendation 8: Utilize all of the databases available to the Transportation Security Operations Center when running passengers who rise to the level of a LEO referral against intelligence and criminal databases.

Concur in principle. TSA has convened a working group composed of members of the Office of Security Operations, Office of Chief Counsel, Office of Intelligence, and the Office of Law Enforcement/FAMS to address this recommendation. This group will conduct a feasibility study during FY 2010 to examine if this recommendation can be fully implemented. This study will look at the various authorities, permissions, and limitations of each of the databases or systems cited in this audit. Access to some of the systems, such as Criminal History Record Check (CHRC), requires more justification than a BDO referral. Because some of the databases or systems contain classified information, TSA will also need to adopt a communication strategy to transmit the passenger information back and forth between the BDO and TSOC. TSA will work on a process to collect the passenger information, verify the passenger's identity, through checks of databases, and analyze that information to determine if the passenger is the subject of an investigation and may pose a risk to aviation.

Recommendation 9: Establish a plan that includes objectives, milestones, and timeframes to develop outcome-oriented performance measures to help refine the current methods used by Behavior Detection Officers for identifying individuals who may pose a risk to the aviation system.

Concur. TSA understands the value of outcome-oriented performance measures. However, as noted by GAO, there is difficulty in establishing these measures for a deterrence-based program. Nonetheless, TSA will consult with industry experts to develop outcome-oriented performance measures. TSA will establish a plan that includes objectives, milestones, and timeframes, with an end result of producing outcome-oriented performance measures to help refine the current methods used by BDOs for identifying individuals who may pose a risk to the aviation system.

Recommendation 10: Establish controls to help ensure completeness, accuracy, authorization, and validity of data collected during SPOT screening.

Concur. In March 2010, TSA migrated the SPOT database to TSA's Performance Management Information System. This migration greatly enhances the SPOT program's capabilities, as they relate to completeness, accuracy, authorization, and validity of data collected during SPOT screening. Additional controls have been put in place to address the shortcomings of the previous database which were highlighted by GAO. TSA is also examining a technology solution to allow one-time transcription of all SPOT referral data. This will reduce the possibility of errors due to incorrect transcription from one medium to another.

Recommendation 11: Establish timeframes and milestones for its plan to systematically conduct evaluations of the SPOT training program on a periodic basis.

Concur. DHS S&T, in conjunction with TSA has sponsored a BDO Job Task Analysis (JTA). Outputs of the JTA will include Knowledge, Skills, Abilities, and Other characteristics of BDOs and Training Learning Objectives. These two items will enable TSA to conduct an in-depth training gap analysis. This analysis will begin immediately following completion of the JTA and will take approximately three months to complete. Upon completion of the training gap analysis, TSA will develop detailed project plans with milestones and schedules based on the scope of the overall curriculum development/revision effort.

Sincerely yours,

Jerald E. Levine
Director
DHS GAO/OIG Liaison Office

Appendix III: GAO Contacts and Staff Acknowledgments

GAO Contact

Stephen M. Lord, (202) 512-4379, or lords@gao.gov

Staff Acknowledgments

In addition to the contact named above, David M. Bruno, Assistant Director, and Jonathan R. Tumin, managed this assignment. Ryan Consaul, Jeff C. Jensen, Kevin Remondini, and Julie E. Silvers made significant contributions to the work. Arthur James, Jr., Amanda Miller, and Douglas Sloane assisted with design, methodology, and data analysis. Chris Dionis assisted with issues related to training. Katherine Davis and Debra Sebastian provided assistance in report preparation; Tracey King and Tom Lombardi provided legal support; and Pille Anvelt and Barbara Hills developed the report graphics.

(440886)

GAO's Mission	The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.
Obtaining Copies of GAO Reports and Testimony	The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday afternoon, GAO posts on its Web site newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to www.gao.gov and select "E-mail Updates."
Order by Phone	The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's Web site, http://www.gao.gov/ordering.htm. Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537. Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.
To Report Fraud, Waste, and Abuse in Federal Programs	Contact: Web site: www.gao.gov/fraudnet/fraudnet.htm E-mail: fraudnet@gao.gov Automated answering system: (800) 424-5454 or (202) 512-7470
Congressional Relations	Ralph Dawn, Managing Director, dawnr@gao.gov, (202) 512-4400 U.S. Government Accountability Office, 441 G Street NW, Room 7125 Washington, DC 20548
Public Affairs	Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800 U.S. Government Accountability Office, 441 G Street NW, Room 7149 Washington, DC 20548

Please Print on Recycled Paper

229

○