NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

STRATEGIC IMPROVEMENTS TO TSA SPOT PROGRAM

by

Brent A. Cotton

March 2015

Thesis Co-Advisors: Lauren Wollman
John Rollins

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 2015	Master's Thesis

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
STRATEGIC IMPROVEMENTS TO TSA SPOT PROGRAM	
6. AUTHOR(S) Brent A. Cotton	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution is unlimited	A

13. ABSTRACT (maximum 200 words)

This study of Transportation Security Administration's Screening Passengers by Observation Techniques (SPOT) program analyzes the Government Accountability Office (GAO) and Office of Inspector General (OIG) recommendations for improvement, as well as strengths and weaknesses of the program not specifically addressed in previous assessments. Any analysis of SPOT must be robust, as it represents one of the few threat agnostic countermeasures not limited by technology and finite detection capabilities.

The GAO has recommended Congress withhold funding from SPOT until further evidence of effectiveness can be produced. The first portion of this study revisits the analysis of GAO and OIG in their respective reports. The GAO audits rely on meta-analyses that suggest human lie detection is no more successful than flipping a coin. This study assesses those claims, and reveals some contextual and analytical limitations of the claims. The OIG report offers similar claims, but adds additional insight into critical strategic areas.

The second portion of this study focuses on the strengths and weaknesses of the SPOT program, including an analysis of several GAO and OIG conclusions. Many of the recommendations were operational in nature, and provided little strategic direction to improve the relevance, effectiveness, and credibility of the program. Analyzing strengths and weaknesses provides insight into more strategic recommendations that may improve the security value of SPOT.

14. SUBJECT TERMS	15. NUMBER OF PAGES
aviation security, behavior detection, Transportation Security Administration, human lie detection, behavioral psychology, SPOT, deception detection, risk-based security	113
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UU

NSN 7540–01-280-5500

Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

STRATEGIC IMPROVEMENTS TO TSA SPOT PROGRAM

Brent A. Cotton
Transportation Security Specialist, Transportation Security Administration, Arlington, VA
B.S., University of West Florida, 1997

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF ARTS IN SECURITY STUDIES (HOMELAND SECURITY AND DEFENSE)

from the

NAVAL POSTGRADUATE SCHOOL
March 2015

Author: Brent A. Cotton

Approved by: Lauren Wollman
Thesis Co-Advisor

John Rollins
Thesis Co-Advisor

Mohammed Hafez
Chair, Department of National Security Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This study of Transportation Security Administration's Screening Passengers by Observation Techniques (SPOT) program analyzes the Government Accountability Office (GAO) and Office of Inspector General (OIG) recommendations for improvement, as well as strengths and weaknesses of the program not specifically addressed in previous assessments. Any analysis of SPOT must be robust, as it represents one of the few threat agnostic countermeasures not limited by technology and finite detection capabilities.

The GAO has recommended Congress withhold funding from SPOT until further evidence of effectiveness can be produced. The first portion of this study revisits the analysis of GAO and OIG in their respective reports. The GAO audits rely on meta-analyses that suggest human lie detection is no more successful than flipping a coin. This study assesses those claims, and reveals some contextual and analytical limitations of the claims. The OIG report offers similar claims, but adds additional insight into critical strategic areas.

The second portion of this study focuses on the strengths and weaknesses of the SPOT program, including an analysis of several GAO and OIG conclusions. Many of the recommendations were operational in nature, and provided little strategic direction to improve the relevance, effectiveness, and credibility of the program. Analyzing strengths and weaknesses provides insight into more strategic recommendations that may improve the security value of SPOT.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
A. PROBLEM STATEMENT 2
B. LITERATURE REVIEW 6
1. Scientific Research 7
a. Emotion 8
b. Intuition 10
c. Human Lie Detection 12
d. Non-verbal Communication 13
2. Official Government Documentation 16
a. Audits/Reports 16
b. Congressional Testimony 20
c. Public Law 22
3. Best Practices for TSA Behavior Detection 22
4. Summary 23
C. RESEARCH DESIGN 24
D. OVERVIEW OF REMAINING CHAPTERS 27

II. BACKGROUND 29
A. HISTORY 29
B. CURRENT STATE 33
1. Concept of Operations 34

III. ANALYSIS OF GAO CLAIMS 37
A. SPOT HAS NOT DEMONSTRATED EFFECTIVENESS 37
B. SPOT HAS NOT CAUGHT A SINGLE TERRORIST 39
C. SPOT FAILED TO INTERCEPT TERRORISTS ON 23 OCCASIONS 41
D. SPOT IS APPLIED INCONSISTENTLY 41
E. LITERATURE DOES NOT VALIDATE THE PRACTICE 44
F. THE IMPORTANCE OF CONTEXT IN DATA ANALYSIS (EVIDENCE) 48

IV. STRENGTHS AND WEAKNESSES OF BEHAVIOR DETECTION 51
A. STRENGTHS 51
1. Removing Criminals from the Transportation System 51
2. Flexibility 52
3. Deterrence Value 53
4. Threat Agnosticism 55
5. Unintended Benefits 56
B. WEAKNESSES 57
1. Utilization 58
2. BDO Selection Criteria 59
3. Training 60
a. Training Inconsistency 60

b. Inadequacy of Training Content 61
4. Effectiveness and Performance Testing..... 62
V. RECOMMENDATIONS..... 65
A. UTILIZATION RECOMMENDATIONS..... 65
B. TRAINING RECOMMENDATIONS..... 71
C. EFFECTIVENESS AND PERFORMANCE TESTING:..... 78
VI. CONCLUSION..... 83
LIST OF REFERENCES..... 87
INITIAL DISTRIBUTION LIST..... 93

LIST OF FIGURES

Figure 1. TSA Layers of Security 3
Figure 2. James Russell's "Conclusions from Recent Reviews on the Universality Thesis" 10
Figure 3. TSA SPOT Referral/Arrest Rates 39
Figure 4. TSA Referral Rates among49 Airports 43
Figure 5. Six Universal Emotions 74

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ASVAB	Armed Forces Vocational Aptitude Battery
ATF	Bureau of Alcohol, Tobacco, Firearms, and Explosives
ATSA	Aviation Transportation Security Act
BASS	Behavior Awareness Screening System
BDO	Behavior Detection Officer
CCTV	Close Captioned Television
CIA	Central Intelligence Agency
CONOPS	concept of operations
CPR	cardiac pulmonary resuscitation
DEN	Denver International Airport
DHS	Department of Homeland Security
EEO	equal employment opportunity
ETD	explosives trace detection
FAM	federal air marshall
FBI	Federal Bureau of Investigation
FLETC	federal law enforcement training center
FTE	full time equivalent
FY	fiscal year
GAO	Government Accountability Office
GED	general education development test
IED	improvised explosive device
KST	known or suspected terrorist
LAX	Los Angeles International Airport
LEO	law enforcement officer
MI	managed inclusion
MSP	Minneapolis St. Paul International Airport
OIG	Office of Inspector General
OJT	on-the-job training
OTWE	Office of Training and Workforce Engagement
P(d)	probability of detection

SIDA	secure identification display area
SME	subject matter expert
SPOT	screening passengers by observation techniques
TAC	threat assessment capabilities
TAD	threat assessment division
TDC	travel document checker
TSA	Transportation Security Administration
TSO	transportation security officer
TSOC	transportation security operations center
U.S.	United States
USSS	United States Secret Service
VIPR	visible intermodal preventative and response team
WTMD	walk through metal detector

EXECUTIVE SUMMARY

The Transportation Security Administration's (TSA) Screening Passengers by Observation Techniques (SPOT) program is under intense scrutiny by Congress, and in danger of losing funding. Recent reports by the Government Accountability Office (GAO) and the Department of Homeland Security (DHS) Office of Inspector General (OIG) have identified several areas for operational improvement, but the overarching question that needs to be answered is whether the program adds security value. The TSA must prove the program is effective before it will receive additional funding for the program. The practice of using behavioral cues to identify criminals or terrorists is in regular practice by many law enforcement and government agencies, but none is as visible as TSA's dedicated workforce operating in a public airport environment. This study examines the literature available on deception detection, as well as the strengths and weaknesses of SPOT to determine if the program should continue to be funded.

To date, several congressional hearings have been held on the topic, to include testimonies from the GAO, DHS OIG, DHS Behavioral Sciences Division, and TSA. GAO and DHS/TSA disagree as to whether the techniques work. Both are able to cite relevant academic and scientific literature supporting their respective arguments, which places the debate firmly in gridlock. Unfortunately, discussions in these "official" environments (hearings, etc.) tend to serve only as a forum for each party to state their case and "double down" on their original stance, which results in little progress being achieved toward a consensus.

An examination of the literature reveals that this gridlock is not unique to the GAO and TSA. The academic community is also divided as to whether deception detection techniques are viable. The division is based on the debate of whether lab studies using trivial lies and unmotivated liars are indicative of real-life performance. The studies are also focused specifically on lie detection, whereas SPOT uses a sequential combination of observation, situational

awareness, logical analysis, and deception detection techniques to determine if a person requires additional screening. The decision is based on the *totality* of circumstances, rather than a single structured lie detection interview. Only a single study, conducted by the DSH Behavioral Science Division, used the entirety of SPOT techniques, and the results indicated that SPOT is nine times, or 900 percent more effective than randomly selecting individuals.[1]

Areas in which the literature is in general agreement are: 1) verbal and non-verbal cues to deception do exist, 2) no "Pinocchio's nose" telltale indicator of deception exists, 3) deception can be easier or more difficult to detect depending on the skill of the liar, 4) high-stakes lies may be easier to detect than trivial lies due to the powerful emotions associated with a motivated lie, but are not simply escalated versions of traditional cues, 5) cues to deception may be more evident during personal lies, 6) lie catchers can be trained to elicit indicators from liars by increasing their cognitive load, and 7) more research in the field is necessary. Particularly, additional research in the field of high-stakes lies is necessary, as it is in its infancy compared to general laboratory deception detection using trivial lies.[2]

Examination of the strengths and weaknesses of SPOT reveals that although SPOT has removed many criminals and dangerous prohibited items from the "transportation system," areas exist in which strategic improvement can improve security value and efficiency of the program. SPOT has the benefit of being flexible, as the officers are a countermeasure that can be rapidly deployed to a variety of locations and activities. Additionally, SPOT is threat agnostic, and is designed to detect an individual with malicious intent, versus a specific, limited

---

[1] *Behavioral Science and Security: Evaluating TSA's SPOT (Screening of Passengers by Observational Techniques) Program: U.S. House of Representatives, Committee on Science and Technology Subcommittee on Investigations and Oversight*, 111th Cong. (2011) (statement of Larry Willis), http://www.dhs.gov/news/2011/04/05/testimony-mr-larry-willis-program-manager-science-and-technology-directorate.

[2] Stephen Porter and Leanne Brinke, "The Truth About Lies: What Works in Detecting High-stakes Deception?," *Legal and Criminological Psychology* 15, no. 1 (2010): 60, accessed September 3, 2014, doi: 10.1348/135532509X433151.

set of explosive types or weapons. SPOT also provides a great deterrent to U.S. adversaries, as it is difficult for an adversary to understand what will happen during an interaction with a SPOT officer, and how to counteract it. Despite these positive attributes, no efforts have been made since SPOT's implementation to improve the program's strategic direction or identify and implement process improvements. The GAO and OIG noted many of SPOT program issues as operational improvements, but strategic improvements can also increase SPOT's security value, efficiency, and accountability. Examination of the documentation, as well as audit reports, reveals that SPOT has poor selection and hiring practices, inconsistent training and insufficient training, and a lack of accountability metrics.

This study concludes that the literature regarding deception detection is limited concerning the ability to reproduce true high-stakes conditions in a laboratory study, and therefore, is not directly applicable to SPOT or other deception detection countermeasures. While the lab studies are a good first step, more research is needed to inform high-stakes deception detection, particularly in the context of terrorism. It is also not advisable to base program funding recommendations on meta-analyses conducted using trivial lies and psychology students as participants. Additional research is also needed in the area of "truth wizards" to determine what psychometric (or other) attributes are relevant to a successful deception detection practitioner.

In the area of strategic improvements to the existing program, this study recommends that the TSA invest in substantial analysis and commit to revising the program based on the best available information. This study concludes that the TSA should undertake a strategic revamping of the program to include the following.

- Establish an operational baseline performance metric for existing behavior detection officers (BDOs) using arrests/prohibited item-to-referral ratio.
- Develop hiring criteria based on a study of psychometric and other attributes of high performing BDOs.

- Place additional hiring emphasis on candidates possessing program-enhancing characteristics, such as language skills and cultural competency/background.

- Conduct all SPOT training at the federal law enforcement training center (FLETC) using established procedures.

- Revising SPOT curriculum to include an explanation of TSA authority, cultural, political, and socioeconomic variables that affect a person's behavior, and how personal biases affect response to those variables.

- Offering advanced training classes in areas that will add value at the checkpoint, as well as offering career advancement opportunities for the BDOs (by collecting and demonstrating proficiency in advanced training areas).

- Collaborating with agencies' training spy craft or undercover techniques to test the BDOs covertly or overtly.

While academia continues to advance the precision and relevance in which deception detection is studied and assessed, SPOT can make incremental gains in effectiveness, efficiency, and understanding of performance, even in lieu of established academic performance indicators. These strategic recommendations will provide valuable improvements and defensibility to SPOT, as will continued evolution of the program based on relevant research. In fact, the TSA has the opportunity to be a contributor to the body of scientific research that exists on the topic.

ACKNOWLEDGMENTS

Attending graduate school after a 17-year hiatus from higher education is challenging to say the least. If not for the support of my wife, Laurie, it may not have been possible. In fact, as I am writing this, I am realizing how many sacrifices she made during the past 18 months, and how much I owe her.

I also need to offer a sincere “thank you” to my organization, the Transportation Security Agency (TSA), for supporting me in this program, and specifically, Assistant Administrators Robin Kane, John Sanders, and Chris McLaughlin for their sponsorship. I hope I will be able to return this investment and more to the agency and the missions we support. Previous graduate Anthony Perry and colleagues Kriste Jordan-Smith and Leigh Otey were also instrumental in motivating me to apply; their encouragement is greatly appreciated.

Lauren Wollman, and John Rollins, who advised on this paper, were great supporters of mine (and all the students) and provide the encouragement, guidance, and counsel needed to complete the program. I must mention that Lauren is a master of her craft. I do not think there could be a better combination, and I appreciate them coping with my casual writing style. I cannot say enough about the professors and staff at the Center for Homeland Defense and Security (CHDS). I will not name them all individually, but I will say that their passion for educating and learning makes the whole journey worthwhile. I appreciate the vision of the Department of the Navy to partner with the CHDS to create this unmatched educational experience. Lastly, I acknowledge my fellow classmates, from whom I have learned so much.

I hope this thesis provides some insight and understanding into the TSA SPOT program, its unique mission, and the challenges the program faces. The recommendations herein are intended to provide logical, defensible, strategic

direction to ensure the program can continue to provide a valuable threat agnostic layer to the aviation system.

I. INTRODUCTION

On December 14, 1999, an Islamic terrorist named Ahmed Ressem attempted to enter the United States from the Canadian border with a trunk full of explosives. The explosive materials were intended for a terrorist attack against the Los Angeles International Airport (LAX), but the attack would never occur. While crossing over from Victoria British Columbia to Port Angeles, Washington, a United States (U.S.) immigration inspector noticed that Ressem's response to his standard questions did not make sense. He stated he was going to Seattle for a 2-day business trip. The response was peculiar for two reasons, 1) the inspector had not asked him why he was going to Seattle, and 2) it would not make sense for a non-tourist to take this route to Seattle. Something was not right. Wants and warrants returned nothing, Ressem's baggage was cleared, and he was sent on his way. At the arrival end of the ferry, Ressem was again subject to routine interaction with another U.S. immigration inspector. This time the inspector noticed him fidgeting, acting jittery, and sweating during routine questioning. This inspector also knew something was not right, and ultimately, searched the car and found explosives intended for the attack on LAX. While several attacks have been thwarted through investigations, this one stands out as being thwarted in progress using existing countermeasures. While intelligence and investigating has provided the greatest anti-terrorism success, it can fall short in the area of homegrown violent extremists or sympathizers not directly affiliated with a known terrorist group. The shift of the terrorist landscape to more one-off attacks or plots reminds everyone of the importance of real-time countermeasures. The countermeasure in this case was a routine passenger engagement by inspectors, who have been trained to detect unusual behaviors in the context of their post. Both inspectors noticed the unusual behavior of Ressem, such as his inability to provide consistent explanations for his travel, his nervousness, twitching, and sweating; behaviors that can be linked to a "fear of discovery" by criminals or terrorists. Their suspicion led to additional screening

and the discovery of the explosives that ultimately thwarted the planned attack on LAX. Real time threat assessment by the officers (using principles of behavior detection) is responsible for the only terrorist attack thwarted in progress by a U.S. government countermeasure. Ressem attempted to cover all his bases, including a cover story, and fake ID that did not arouse suspicion; however, he was not able to suppress his fear of discovery.

A. PROBLEM STATEMENT

The Transportation Security Administration's (TSA) Screening Passengers by Observation Techniques (SPOT) program was established to create a threat agnostic, non-technology based countermeasure capable of identifying a malicious person rather than a specific type of threat item, such as an improvised explosive device (IED), a gun, or other threat object. Being able to identify terrorists by their pre-event behavior, regardless of the threat type or event planned, breaks the cycle of traditional, reactive game theoretic responses. Without threat agnostic layers, the TSA would solely rely on technology-based layers that are only effective against a specific type of threat. For example, red team (adversary) attempts to use a shoe IED against an aircraft, blue team (TSA) X-rays all shoes or red team uses IEDs concealed on a person, blue team purchases and implements body scanners. This reactive cycle typically keeps the blue team guessing, and at risk of a new style of attack. SPOT was implemented to give the TSA a real-time, threat-agnostic layer capable of identifying persons with malintent, regardless of what type of threat—explosive or contraband—they may be carrying. Game theorists would call it "disruptive technology" and operates outside the parameters of the "game." With the exception of the important intelligence-based layers (which are not necessarily real time), the TSA layers listed in Figure 1 include mainly countermeasures designed to detect things versus terrorists.

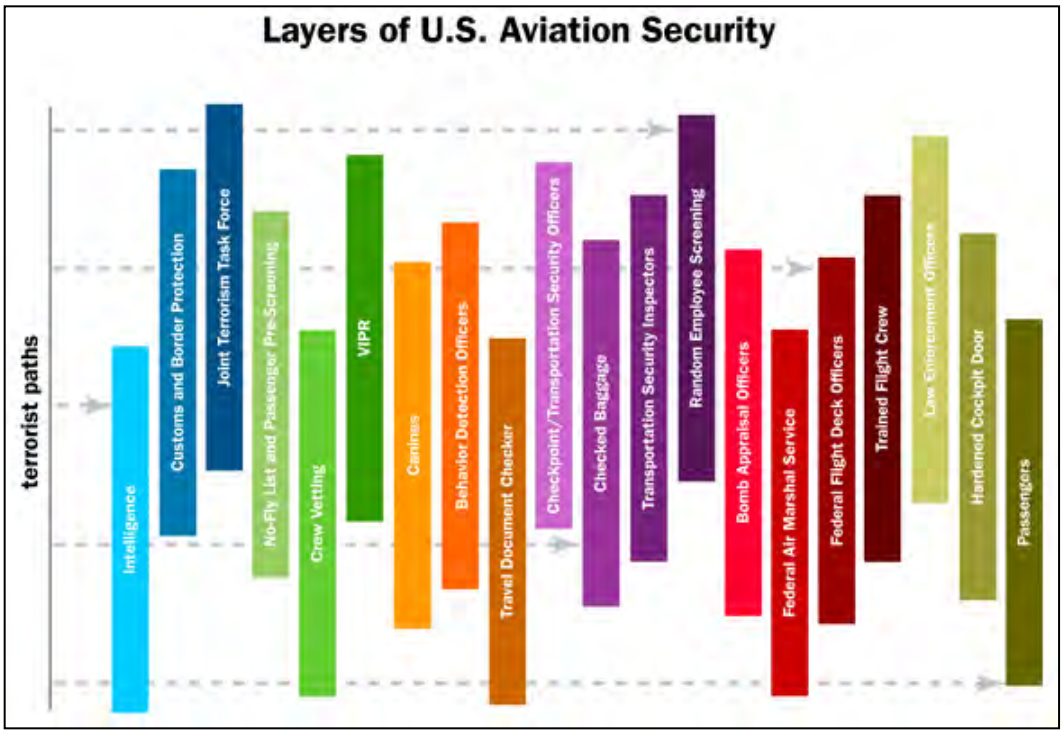

Figure 1. TSA Layers of Security[1]

Although the SPOT program is the only real time, threat agnostic layer in TSA's layers of countermeasures, it is currently in jeopardy of being cut by Congress, or at a minimum, receiving limited or reduced funding. In the last two years, the Government Accountability Office (GAO) and the Office of the Inspector General (OIG) have criticized TSA's SPOT program for lack of effectiveness evidence, poor training consistency, operational issues, and a lack of strategy. These criticisms correlate to two major areas, a perceived gap in the validity of behavior detection, and the development and management of the program. Many of the criticisms were valid operational concerns, and are already being addressed in response to GAO's and OIG's recommendations. For example, improving data entry after an incident is a valid operational criticism, which is easily addressed and will undeniably improve the program. While an important improvement, it does not impact the program's future direction,

[1] "Layers of Security," last modified July 23, 2014, http://www.tsa.gov/about-tsa/layers-security.

effectiveness, or credibility, but rather falls in the category of operational improvement versus strategic game-changer. During this period of scrutiny, the TSA should endeavor to make real strategic enhancements to the effectiveness of behavior detection in an aviation security context. With nearly 3,000 behavior detection officers (BDOS) and an approximate $200 million annual budget, the TSA owes it to the taxpayers to ensure SPOT is continually improving and providing added value to transportation security.

Contributing to the GAO's criticisms are the general lack of concurrence amongst behavioral science professionals, allegations of profiling, lack of multitasking ability, and general lack of knowledge about the current threat environment and the value of deterrence. The GAO audits contend that not enough is known about the science of behavior detection to assert its effectiveness; however, little disagreement exists that people will exhibit signs of fear or anxiety during times of elevated stress. The disagreement is more about whether these signs are consistently detectable through interaction with the subject. A number of scientific studies and practical publications support the validity of the practice, including an independent evaluation by the Department of Homeland Security (DHS) that found the practice to be much more effective than random selection.

Many federal and law enforcement agencies, such as the Central Intelligence Agency (CIA), Federal Bureau of Investigation (FBI), and the United States Secret Service (USSS), have been utilizing these concepts for years. The principles for the TSA version of behavior detection are based on the Israeli model, as well as research from Dr. Paul Ekman, who pioneered the study of non-verbal communication as it relates to deception detection. Those who practice SPOT or similar programs are sometimes referred to as human lie detectors; a bit parochial and oversimplified, but somewhat accurate. The use of behavior detection as a security countermeasure is based on the generally accepted universality of the seven defined human emotions. Many of the criticisms of SPOT seem to focus on a single aspect, such as human lie

detection through the recognition of micro-facial expressions, or other singular, controversial practice. Research about the program reveals that trained practitioners actually use a cadre of techniques to assess the totality of the existing circumstances. Practitioners use a combination of observation, casual conversation, directed conversation, and response evaluation (whether verbal or physiological) to evaluate whether a passenger is being deceptive, or is nervous, fearful, or uncomfortable for some other reason.

The TSA still has to solve challenging problems, including how to begin measuring effectiveness for the program given that no actual terrorists have been caught as a result of SPOT. The only data available is the number of persons arrested at the airport for doing or possessing something illegal. Many success stories in this area has occurred; however, some argue that using arrest data only amounts to proxy data, and may not translate to success for catching a terrorist. It is also difficult to translate the number of criminal “catches” into effectiveness since it is not known how many were missed. For example, if 10 drug mules are identified at an airport using behavior analysis, and 10 get through without being discovered, then there the probability of detection (P(d)) is 50 percent. However, in reality it is not possible to know about the 10 that were missed Thus, it is also not possible to assign a P(d) or success rate confidently. It can be argued that since no aviation attacks have occurred on U.S. soil, then no terrorists with malintent have been missed. To that end, the largest unknown may be how to measure deterrence, and whether deterrence is responsible for the lack of terrorist attacks on the U.S. civil aviation system.

The TSA is shifting to a layered, risk-based security approach, with an increase in reliance on intelligence information, and subsequently, applying technology resources strategically to have the most value in guarding against a potential attack. This shift is in part due to the acknowledgement that the federal government cannot develop and buy a new “scanner” for every type of potential threat, and therefore, it makes sense to provide additional focus on who is flying, not just what they are carrying. Without the real-time layer of screening for

"intent," the traveling public will be reliant on a security system that relies heavily on automated technologies, which are limited to a finite (and potentially documented) detection capability. Due to the number of variables involved and the limitations of lab testing, it is not likely that the scientific community will soon reach a consensus about the effectiveness of behavior detection techniques. However, in lieu of a proven scientific method to measure and quantify the P(d), this study will attempt to determine if ways exist to improve the program based on the literature available, assess the GAO conclusions and analysis, determine if methods exist to begin measuring effectiveness using existing data, and offer additional strategic improvements to make SPOT viable and credible in a large-scale aviation security context.

B. LITERATURE REVIEW

The science of behavior detection as an anti-terrorism technique is not well documented, yet many anti-terrorism organizations continue to practice and train in these techniques. State and local law enforcement agencies also train and practice similar techniques without criticism or question. The documentation that does exist is generally only focused on one aspect of behavior detection, such as human lie detection or non-verbal deception, but never the topic as a whole. Much of the existing documentation is controversial and can essentially be divided into two distinct opinions, recognized experts who validate the effectiveness of such practices, and those who believe the success of such practices is a myth. While the academic studies have mixed results, both supporters and detractors agree that the results are likely impacted by the inability to recreate conditions and genuine emotions, such as fear, stress, and anxiety associated with deception. The level of fear, stress, or anxiety is directly proportional to the consequence of being discovered. A trained observer can visually detect this fear or anxiety as it manifests through involuntary physical and/or physiological activity, including but not limited to increased heart rate, facial displays of emotion, changes in speed and direction of movement, or

nervous sweating.[2] This inability to elicit genuine emotion during lab testing is not an insignificant limitation, as behavior detection techniques rely heavily on the indications of basic emotions.

A large portion of the research around deception detection is primarily focused on "human lie detection" (quite different from SPOT) and consists primarily of academic publications, such as psychology textbooks, journals, and academic theses; but the research also includes legislation, congressional reports, news articles, and expert opinion. This literature review examines the scientific research on the topic, official government documentation, and best practices that may improve the state of the program.

1. Scientific Research

Research of SPOT reveals that it is not singularly based on the ability to determine if a person is lying, but rather on the assessment of the totality of the existing circumstances including a person's appearance (not related to race, religion, etc.), behavior, potential responses to questioning or casual conversation, and the validity of their responses compared with their travel plans and travel documentation. Only a single scientific study specifically about TSA SPOT exists and was an independent evaluation ordered by the DHS to determine the effectiveness of the practice (versus the science behind the practice); this study is discussed later under "official government documentation." As no scientific studies cover the entire breadth of SPOT techniques, it is necessary to evaluate scientific evidence in similar areas, such as the universality of human emotion, intuition, human lie detection, and non-verbal communication. While these topics do overlap, and none of them independently could be said to validate or invalidate SPOT, the totality of their results may help inform the validity of the practice of SPOT.

[2] *The TSA Spot Program: A Law Enforcement Perspective: Hearing before the U.S. House of Representatives Committee on Science, Space, and Technology Subcommittee on Investigations and Oversight*, 111th Cong., 5 (2011) (statement of Detective Lieutenant Peter J. DiDomenica).

a. Emotion

Determining whether people all portray the same indicators of emotion is important basis for evaluating a human being's response to different circumstances. For example, if what westerners call "fear" manifested itself as happiness in different cultures, then observers would constantly be mistaking the intentions or responses of other cultures. According to psychologist Dr. David Matsumoto:

> Emotions are not just feelings. Emotion…..is one class of affective phenomenon. To me, emotions are transient, bio-psycho-social reactions designed to aid individuals in adapting to and coping with events that have implications for survival and well being. They are biological because they involve physiological responses from the nervous system, and prime skeletal muscle activities. They are psychological because they involve specific mental processes required for elicitation and regulation of response. And they are social because they are often elicited by social interactions, and have meaning with those interactions.[3]

Psychologists generally agree that all people experience a few basic emotions, regardless of cultural background or geographic location (see Figure 2). Seven commonly accepted primary emotions are the following: enjoyment, surprise, fear, anger, contempt, disgust, and sadness as defined by Matsumoto.[4] Psychologists debate whether five, six, or seven universal emotions exist, but they only differ by combining or extracting certain emotions. For example, some studies may list fear and surprise as a single expression, for a total of six universal emotions.[5] More specifically, James Russell's review of cross cultural emotion studies claims that modern research acknowledges that the face involuntarily reveals these basic emotions in all humans, regardless of origin.[6] Modern research on the universality of emotion in general was pioneered

---

[3] David Matsumoto, "The Origin of Universal Human Emotions," 1, October 18 2014, http://davidmatsumoto.com/content/NG%20Spain%20Article_2_.pdf.

[4] Ibid., 3.

[5] James Russell, "Is There Universal Recognition of Emotion From Facial Expressions? A Review of the Cross Cultural Studies," *Psychological Bulletin* 115, no. 1 (1994): 102.

[6] Ibid.

by Dr. Ekman's Department of Defense funded non-verbal communication research in the 1960s.[7] Since then, it has been widely accepted amongst psychologists and deception detection researchers (such as Matsumoto, DePaulo, and many others) that universality of emotion does exist, and automatic links between the emotion and facial expressions do occur.[8] Psychologists who dispute the universality of human emotion are typically only referring to the tertiary, or more complex set of emotions not tied to survival, such as love, ambition, pride, self-respect, shame, guilt, inspiration, enthusiasm, sadness, awe, admiration, humility and humiliation, sense of justice and injustice, envy, malice, resentment, cruelty, hatred, etc.[9] These emotions may be culturally specific or unique to certain cultures, particularly the west. The universality of basic or primary emotions and their associated facial expressions is strong support for deception detection being universal across cultures, as the fear of discovery should manifest itself similarly amongst all cultures. For example, an American should be able to detect deceptive behavior from a Japanese person based on the universality of basic emotions, and the fear of discovery associated with deceptive activity, such as terrorist or criminal activity.

---

[7] Paul Ekman, *Emotions Revealed: How Recognizing Faces and Feelings Improves Communication and Emotional Life* (New York: Times Books, 2003), 2.

[8] Russell, "Is There Universal Recognition of Emotion From Facial Expressions? A Review of the Cross Cultural Studies," 102–103.

[9] Liah Greenfeld, "Are Human Emotions Universal?," *Psychology Today: Health, Help, Happiness + Find a Therapist*, April 23, 2013. http://www.psychologytoday.com/blog/the-modern-mind/201304/are-human-emotions-universal.

FACIAL EXPRESSIONS OF EMOTION

Table 1
Conclusions from Recent Reviews of Research on the Universality Thesis

Source	Conclusion
1. Ekman (1980)	"Are facial expressions of emotion the same for all human beings?" (p. 91). "Definitive data are now available on the question of universality" (p. 93). "There is conclusive scientific evidence to resolve the question of universality" (p. 96). "There are some facial expressions of emotion which are universal" (p. 137).
2. Izard (1980)	"Impressive evidence for the innateness and universality of six of the fundamental emotions: enjoyment (happiness), distress (sadness), anger, disgust, surprise, and fear" (p. 201). "Since all human beings recognize these expressions and attribute to them the *same* experiential significance, it is reasonable to infer that they are genetically based or preprogrammed" (p. 185).
3. Frijda (1986)	"Many facial expressions . . . occur throughout the world in every human race and culture. The expressions appear to represent, in every culture, the same emotions" (p. 67).
4. Fridlund, Ekman, and Oster (1987)	"Observers label certain facial expressions of emotion in the same way regardless of culture" (p. 157).
5. Gudykunst and Ting-Toomey (1988)	"The research cited indicates that facial expressions representing the basic emotions are recognized universally" (p. 182).
6. Buck (1988)	"Research on the communication of emotion via facial expression suggests that certain displays appear and are correctly recognized in widely different cultures" (p. 351).
7. Izard and Saxton (1988)	"The evidence for the innateness and universality of the expressions of the fundamental emotions is sufficiently robust to consider Darwin's hypothesis as an established axiom of behavioral science" (pp. 651–652).
8. Oster, Daily, and Goldenthal (1989)	"Conclusive evidence for the universality of certain facial expressions has come from studies in which observers were asked to identify the emotions shown in photographs of facial expressions" (p. 114).
9. D. E. Brown (1991)	"The conclusion seems inescapable: There are universal emotional expressions" (p. 26).
10. Mesquita and Frijda (1992)	"Certain facial expressions of emotion appear to be universal across cultures" (p. 14). "There appears to exist a universal human set of emotion reaction modes [including] facial expressions" (p. 21).
11. Carlson and Hatfield (1992)	"Ekman and other psychologists have uncovered compelling evidence that six basic emotions are expressed in much the same ways in all cultures" (p. 221).

Figure 2. James Russell's "Conclusions from Recent Reviews on the Universality Thesis"[10]

b. Intuition

The theory of "human lie detectors" is as fundamental as a mother's intuition. Intuition may be better described as subconscious observation, perception, and subsequent response. For example, children who have been sneaking cookies are subconsciously displaying "indicators" to their mother that they have done something wrong, and their mother subconsciously knows

[10] Russell, "Is There Universal Recognition of Emotion From Facial Expressions? A Review of the Cross Cultural Studies," 103.

something is different about the child's behavior; thus, the "deception" is "detected." This theory is supported in psychological circles as "listening with your eyes."[11] This capability is enhanced when baseline data is available as a mental comparison, as in the case of a mother and child. Other similar notions, such as being a "good judge of character," or having good intuition, or having a gut feeling, support the validity of capabilities, such as deception detection, although they do not explain it. Recently scientists have begun to study intuition, and Carlin Flora explained it in *Psychology Today*:

> Intuitions, or gut feelings, are sudden, strong judgments whose origin we can't immediately explain. Although they seem to emerge from an obscure inner force, they actually begin with a perception of something outside—a facial expression, a tone of voice, a visual inconsistency so fleeting you're not even aware you noticed. Think of them as rapid cognition or condensed reasoning that takes advantage of the brain's built-in shortcuts. Or think of intuition as an unconscious associative process. Long dismissed as magical or beneath the dignity of science, intuition turns out to muster some fancy and fast mental operations. The best explanation psychologists now offer is that intuition is a mental matching game. The brain takes in a situation, does a very quick search of its files, and then finds its best analogue among the stored sprawl of memories and knowledge. Based on that analogy, you ascribe meaning to the situation in front of you. A doctor might simply glance at a pallid young woman complaining of fatigue and shortness of breath and immediately intuit she suffers from anemia.[12]

Flora continues to reference the research of Dr. Maureen O'Sullivan, which indicates that some people are naturally excellent at detecting deception, and that they have a strong interest in people in common, as well as a broad range of experience to draw upon.[13] Good deception detectors who have broad life experience also supports the notion that the brain is performing an instantaneous

---

[11] Malcolm Gladwell, *Blink: The Power of Thinking without Thinking* (New York, NY: Back Bay Books, 2007), 245–254.

[12] Carlin Flora, "When to Go With Your Gut," *Psychology Today*, May 1, 2007, www.psychologytoday.com%2Fcollections%2F201205%2Fgo-your-gut%2Fyour-best-guide.

[13] Ibid.

analysis based on the totality of circumstances, and matching those circumstances against other experiences and outcomes.

c. Human Lie Detection

"Decades of research has shown that people are poor at detecting lies."[14] However, despite decades of research to the contrary, a plethora of books claim to teach how to become a human lie detector, many written by psychologists, former spies, or government agents. It has been indicated that the science is used and accepted by practitioners and operators, as is the number of legitimate agencies and organizations that have resources dedicated to deception detection. Of course, some academic research supports the notion of human lie detectors, and much of the literature that acknowledges these capabilities, also claims some persons have a predisposed ability to catch liars, which indicates it may be possible to choose candidates with those characteristics.[15] Other studies question whether these indicators even exist, and claim that the effectiveness of human deception detection is barely better than a coin flip.[16] One comprehensive meta-analysis by Hartwig and Bond Jr. set out to determine "why lie catchers fail" by offering the hypothesis that they rely on inaccurate cues to deception, but instead, they found that lie catchers, when successful, were actually relying on intuitive and accurate cues to deception, even though they were not able to articulate initially why they caught the liar. The conclusion was that lie catchers do not need training on objective cues to deception, as they are intuitively using

---

[14] Maria Hartwig and Charles F. Bond, "Why Do Lie-catchers Fail? A Lens Model Meta-analysis of Human Lie Judgments," *Psychological Bulletin* 137, no. 4 (2011): 643, doi: 10.1037/a0023589.

[15] Mark G. Frank, Melissa A. Menasco, and Maureen O'Sullivan, "Human Behavior and Deception Detection," in *Handbook of Science and Technology for Homeland Security*, vol. 5 (John Wiley and Sons, 2008), 5, sec. 2, accessed September 1, 2013, http://www.cl.cam.ac.uk/~rja14/shb10/frank2.pdf.

[16] United States Government Accountability Office, *Testimony Before the Subcommittee on Transportation Security, Committee on Homeland Security, House of Representatives Aviation Security: TSA Should Limit Future Funding for Behavior Detection, Statement of Stephen M. Lord, Director Homeland Security and Justice* (GAO-14-158T) (Washington, DC: U.S. Government Accountability Office, 2013), 3.

them already; but would benefit from training on techniques to elicit valid cues to deception from the liar.[17]

d. Non-verbal Communication

"The face is a dynamic canvas on which emotions and intentions are communicated, and is scrutinized during all social interactions."[18] Aldert Vrij published *Detecting Lies and Deceit: Pitfalls and Opportunities,* which is widely regarded as the most comprehensive study of deception detection. In this book, Vrij offers an explanation of why deception detection is so difficult, and explains why non-verbal cues to deception are dependent on the liar's personality and the situation. Vrij's explanation includes Zuckerman, DePaulo, and Rosenthal's multi-factor model, which indicates a liar's non-verbal behavior is dependent on whether the liar is emotional, under a high cognitive load (trying to keep his story straight), or attempting to control his behaviors.[19] For example, an emotional liar (such as a lone-wolf terrorist attempting to conduct an attack) may exhibit traditionally accepted signs of deception, such as nervousness, gaze aversion, fidgeting, speech errors, voice pitch change, etc., that are associated with emotions, such as fear, guilt, and excitement.[20] A liar under high cognitive load (such as an Al-Qaeda foot soldier of moderate capability carrying out an order, using a rehearsed story) may exhibit non-verbal indicators, such as thinking hard, slow responses, slow speech, limited detail, and long pauses, which are associated with the extra mental effort required to remember, tell, and defend a rehearsed lie.[21] The third type of liar, individuals attempting to control their behaviors (such as a skilled insider threat with knowledge of the system) may

---

[17] Hartiwig and Bond, "Why Do Lie-catchers Fail? A Lens Model Meta-analysis of Human Lie Judgments," 655–656.

[18] Stephen Porter and Leanne Brinke, "The Truth About Lies: What Works in Detecting High-stakes Deception?," *Legal and Criminological Psychology* 15, no. 1 (2010): 65, accessed September 3, 2014, doi: 10.1348/135532509X433151.

[19] Aldert Vrij, *Detecting Lies and Deceit: Pitfalls and Opportunities* (Chichester: John Wiley, 2008), 38.

[20] Ibid., 39.

[21] Ibid., 39–40.

exhibit indicators, such as overly focused gaze (eye contact) and lack of normal body movements, which makes them appear rigid, rehearsed, or lacking normal spontaneity.[22] Vrij introduces the additional variable of the "liar's personality," which indicate that personality and skill can make liars less susceptible to fear, guilt, or excitement, and the associated indicators, making deception detection more difficult.[23] Vrij's book appears to offer an explanation as to why decades of research have yielded such varying results. Essentially, many of the traditional techniques and indicators have merit, but only if the lie catcher has the ability to categorize the "liar," and only if the lie catcher is better than the liar.

Additionally, Dr. Paul Ekman, who advised on the TSA SPOT program startup, has pioneered scientific research supporting the validity of deception detection. Dr. Ekman's material contains extensive research into non-verbal communication, and test results with thousands of data points to include photographs and video of test subjects.[24] It is widely acknowledged and documented in both the psychology and law enforcement communities that general, detectable behaviors do indicate deception. However, respected psychologists are divided as to whether humans can successfully detect and identify those behaviors. While it is well documented that cues of deception are weak at best during low stakes laboratory experiments, Porter and Brinke note that "facial expressions are of great relevance in betraying motivated lies because of the difficulty of faking genuine emotions and the involuntary nature of the expression of powerful concealed ones."[25] However, they also warn of over reliance on certain indicators, rather than relying on the totality of circumstances and the deviation from baseline.

---

[22] Vrij, *Detecting Lies and Deceit: Pitfalls and Opportunities*, 41–43.

[23] Ibid., 43–46.

[24] Paul Ekman, "Lie Catching and Microexpressions," in *The Philosophy of Deception*, ed. Clancy W. Martin (Oxford: Oxford University Press, 2009).

[25] Porter and Brinke, "The Truth About Lies: What Works in Detecting High-stakes Deception?," 65.

A consistent criticism of lab research in this area is that the methodology of using mostly college students in laboratory experiments limits the validity of the study, as the students are not motivated or rewarded to lie successfully, and are not free to choose to tell a lie or truth.[26] They simply follow the instructions given to them, which may not produce the same indicators of a real liar. Additionally, psychological studies in general are receiving criticism for using these often teen-aged college students as a representative sample of human behavior worldwide.[27] Dr. John Grohol has cast doubt on the usefulness of the data provided by these limited studies, by stating, "the contribution to our real understanding of human behavior is increasingly being called into question."[28] Additional limitations include the lack of high-stakes lies in deception detection deception studies. DePaulo et al., note that cues to deception are stronger when lies are about transgressions,[29] which indicates that subjects forced to tell a lie of no consequence may not exhibit the same cues to deception as someone with a stake in the lie, which limits the usefulness of the study. A 2009 study by O'Sullivan and colleagues compiled results of deception detection conducted at 31 different police departments, in eight countries, using high and low stakes lies. The high-stakes lies were detected at rate of 67.15 percent, while the low stakes lies were only detected at a rate of 55.17 percent.[30]

Overall, the science of deception detection is well documented by psychologists and practitioners, although conclusive academic study results are difficult to obtain (due to small sample sizes of real incidents and difficulty in recreating genuine emotional reactions in a lab environment) and are divisive

---

[26] Bella M. Depaulo et al., "Cues to Deception," *Psychological Bulletin* 129, no. 1 (2003): 106, doi: 10.1037/0033-2909.129.1.74.

[27] John Grohol, "Psychology Secrets: Most Psychology Studies Are College Student Biased," *Psych Central.com*, accessed November 9, 2014, http://psychcentral.com/blog/archives/2010/08/26/psychology-secrets-most-psychology-studies-are-college-student-biased/.

[28] Ibid.

[29] Depaulo et al., "Cues to Deception," 74.

[30] Maureen O'Sullivan et al., "Police Lie Detection Accuracy: The Effect of Lie Scenario," *Law and Human Behavior* 33, no. 6 (2009): 533, doi: 10.1007/s10979-008-9166-4.

across the range of literature. Furthermore, only general results are repeatable, and the test conditions vary greatly. The literature is in general agreement on several factors of deception detection: 1) verbal and non-verbal cues to deception do exist, 2) no "Pinocchio's nose" telltale indicator of deception exists, 3) deception can be easier or more difficult to detect depending on the skill of the liar, 4) high-stakes lies may be easier to detect than trivial lies due to the powerful emotions associated with a motivated lie, but are not simply escalated versions of traditional cues, 5) cues to deception may be more evident during personal lies, 6) lie catchers can be trained to elicit indicators from liars by increasing their cognitive load, and 7) more research in the field is necessary. Particularly, additional research in the field of high-stakes lies is necessary, as it is in its infancy compared to general laboratory deception detection using trivial lies.[31]

2. Official Government Documentation

Numerous official government documents provide insight into the background, performance, current state, and shortcomings of the behavior detection program, including GAO reports, congressional testimony, OIG reports, and official congressional testimony.

a. Audits/Reports

At least three official inquiries into the validity of the TSA SPOT program include two GAO audits and an OIG investigation. Both inquiries found similar issues with the program and provide a combined total of 17 recommendations for improvement to SPOT.

[31] Porter and Brinke, "The Truth About Lies: What Works in Detecting High-stakes Deception?," 60.

The GAO has identified a number of valid recommendations that can be easily addressed by TSA.[32] The initial 2010 report recommended an independent validation of the SPOT program to ensure it is based on valid scientific principles. A second follow up report was issued in 2013, which recommended limiting funding to SPOT based on the lack of evidence that the program is effective, and rebutted the independent validation study that the DHS conducted. While several of the recommendations are valid operational improvements, some of the conclusions appear to be based on incomplete data analysis or a lack of contextual knowledge about the data. The report recognizes the incomplete data at times, but auditors were forced to draw conclusions from a limited data set provided by the TSA. For example, the GAO notes the inconsistency of arrest rates across the country as suspect; however, it is unknown if more suspicious activity occurs in Boston than in Orlando, which explains the variation. On several occasions, the GAO notes a lack of appropriate data from the TSA, and without a more robust analysis, neither the TSA nor GAO will know the cause of the variation in arrest rates. The report notes several anecdotal success stories of SPOT, but again, the TSA did not have sufficient documentation to allow these stories to be considered in the analysis. The report uses the only data available as conclusive evidence of success, arrest data, when in fact arrest data is only useful in context and not as a singular measure of success (or failure). Furthermore, the report does not consider any measure of deterrence associated with the program. At times in the report, it seems as though the GAO does not believe or acknowledge that the science of behavior analysis is valid, and at other times, it appears it agrees with the science but not TSA's execution of the program. Overall, the reports offer a light data analysis using the limited data provided, but the analysis could be improved with additional data and contextual background. The TSA data

[32] United States Government Accountability Office, *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, DC: U.S. Government Accountability Office, 2010), 60–62, http://www.gao.gov/new.items/d10763.pdf.

analyzed in this thesis is contained within these GAO audit reports as well, and is assumed to be accurate.

The GAO analysis is heavily reliant on a well-known deception detection meta-analysis that claims even the best human lie detectors are only slightly better than a coin toss at picking up the deception.[33] However, as the premise of the science is that fear of discovery and the associated nervousness or anxiety causes uncontrollable physiological reactions or "indicators," it is difficult to have confidence in compiled laboratory test results when it is difficult to reproduce the conditions of a real, motivated lie. In fact, one of the meta-analyses used by the GAO specifically identifies this factor as a limitation with the conclusion, "Perhaps liars in the majority of the laboratory research conducted so far are not facing enough of a challenge to give rise to valid behavioral differences. In most of these studies, people are asked to provide a statement with no risk of being challenged about particular details and no risk of being disproven by external information."[34] This test limitation minimizes the utility of scientific tests in this area due to the inability to induce the type of stress capable of eliciting the desired behavioral indicators artificially.

These meta-analyses are strictly focused on deception detection during a structured interview or simple lie detection, which is not a direct correlation to SPOT, but offers general insight into the field. Unfortunately, limited studies of true high-stakes deception detection are available, and as such, many of the studies in the meta-analysis rely on psychology students as subjects and do not provide any training to either the "liar" or the "detector" in the test. While the meta-analysis itself may be valid for the purpose for which it was designed, it is not directly representative of the SPOT program and the conclusions do not

---

[33] United States Government Accountability Office, *Testimony Before the Subcommittee on Transportation Security, Committee on Homeland Security, House of Representatives Aviation Security: TSA Should Limit Future Funding for Behavior Detection, Statement of Stephen M. Lord, Director Homeland Security and Justice*, 3.

[34] Hartiwig and Bond, "Why Do Lie-catchers Fail? A Lens Model Meta-analysis of Human Lie Judgments," 657.

appear to be sufficient to draw concrete conclusions about SPOT. Other meta-analyses claiming positive test results address some of these limitations by first providing training; they claim that deception detection capability can be improved with training and proper candidate selection.[35] Porter and Binke claim that a two-day intensive training on verbal and non-verbal cues to deception improved a group of parole officers' lie detection capability from 40.4 percent accuracy to 76.7 percent accuracy. The GAO audit reports mentioned but did not expressly address the implication of using meta-analyses or the limitations of the test conditions within.

The OIG also conducted a review of SPOT to determine the efficiency, effectiveness, cost-effectiveness, and objectivity of the program.[36] The results of the OIG investigation produced similar operational concerns to the GAO audit, with a bit more focus on training. Several of their recommendations were related to accountability, and the focus was geared toward effectiveness and efficiency versus scientific validation. While crossover occurred between the two reports, the OIG did raise a particularly insightful topic, the selection of BDO personnel. The TSA lacks credibility in this area, as only transportation security officers (TSOs) are able to apply for the position. Attached to the document is TSA's official response to the investigation, where they concur with most recommendations. Unfortunately, the TSA does not address the BDO selection process in the response.[37] Additional TSA-provided data is contained in this report, used in this thesis, and assumed to be accurate.

---

[35] Julia Shaw, Stephen Porter, and Leanne Ten Brinke, "Catching Liars: Training Mental Health and Legal Professionals to Detect High-stakes Lies," *Journal of Forensic Psychiatry & Psychology* 24, no. 2 (January 17, 2013): 2, doi: 10.1080/14789949.2012.752025.

[36] Office of Inspector General, *Transportation Security Administration's Screening of Passengers by Observation Techniques (Redacted)* (OIG-13-91) (Washington, DC: Department of Homeland Security, 2013), 1.

[37] Ibid., 22–28.

b. Congressional Testimony

These audits and reports resulted in an official Congressional hearing that revealed valuable insights from parties on both sides. Unfortunately, after several audits, and an independent validation study, disagreement still occurs as to whether the program is based on sound principles. Examination of both GAO reports and the associated congressional testimony reveal that the GAO is not against behavior detection, but is rather requiring the TSA to provide proof of effectiveness. In 2011, Mr. Larry Willis discussed the results of an independent validation study of SPOT (commissioned by the DHS). Although the report is not published for public consumption, his sworn testimony discussing the results states that SPOT is nine times, or 900 percent more effective than randomly selecting individuals.[38] The latest GAO report claims the data set used is unreliable and that some of the test methodologies were flawed. The report claims that the database used for data collection can only record eight of the 94 behaviors, six signs of deception, and four types of prohibited items. While true at the time, it does not appear to invalidate the overall results of the study.[39] It may limit the usefulness of the data in determining which of the 94 observed behaviors are most used, but it does not invalidate the conclusion that the BDO-selected population was found to have prohibited items nine times more often than randomly selected population. This conclusion is a point of contention between the GAO and DHS, as the DHS claims GAO's analysis of the SPOT validation study led to misleading conclusions, and the GAO responds, "We disagree with this statement."[40] Although not a topic of this thesis, this never ending cycle seems to indicate that the required method of reviewing and responding to GAO audits in public documented forums is not conducive to

---

[38] *Behavioral Science and Security: Evaluating TSA's SPOT (Screening of Passengers by Observational Techniques) Program: U.S. House of Representatives, Committee on Science and Technology Subcommittee on Investigations and Oversight*.

[39] United States Government Accountability Office, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection*, 4–5.

[40] Ibid., 9.

program improvement, and leads to all parties doubling down on their previously documented opinions.

The literature also contains testimony of TSA Administrator John Pistole defending the validation study, as well as the layered security concept that benefits from threat agnostic, non-technology based counter-measures.[41] Additional congressional testimony by TSA Deputy Administrator John Halinski describes the TSA's move toward risk-based security, including a focus on training at the new TSA Academy operating in partnership with the federal law enforcement training center (FLETC) in Glynco, GA.[42]

Outside sources also support the program, evident in Detective Lieutenant Peter J. DiDomenica's statement before the U.S. House of Representatives Committee on Science, Space, and Technology.[43] Although at times it appears his testimony is self-serving, he does defend the techniques associated with SPOT and makes common sense recommendations for the program. The testimony also highlights that SPOT authority was upheld by the Supreme Court's decision that not all transportation modes are an inherent right of a U.S. citizen, which allow certain aspects of air travel to be regulated. Also of note, Congressional committee and sub-committee websites related to transportation and transportation security are months behind in posting documentation related to public hearings and Congressional testimony.

---

41 *Behavioral Science and Security: Evaluating TSA's SPOT (Screening of Passengers by Observational Techniques) Program: U.S. House of Representatives, Committee on Science and Technology Subcommittee on Investigations and Oversight*, 111th Cong. (2011) (statement of Larry Willis), http://www.dhs.gov/news/2011/04/05/testimony-mr-larry-willis-program-manager-science-and-technology-directorate.

42 *Eleven Years After 9/11 Can TSA Evolve To Meet the Next Terrorist Threat?: Hearing before the U.S. House of Representatives Committee on Homeland Security, Subcommittee on Transportation Security*, 112th Cong. (2012) (testimony of John Halinski).

43 *The TSA Spot Program: A Law Enforcement Perspective: Hearing before the U.S. House of Representatives Committee on Science, Space, and Technology Subcommittee on Investigations and Oversight*.

c. Public Law

To complete the literature review on any aviation security topic, include public law 107–71, the *Aviation Transportation Security Act* (ATSA) must be included.[44] This public law was prompted by and enacted shortly after the 9/11 tragedy and provides the basis for the legal obligations and allowances of the U.S. government as it relates to civil aviation security. The ATSA represents the creation of TSA and designates the authorities for screening U.S. mail, carry-on and checked aviation luggage, as well as aviation passengers. Section 114 of the ATSA grants the TSA the responsibility for security in all modes of transportation. Section 114(f) grants the TSA authority to "receive, assess, and distribute intelligence information related to transportation security" and "assess threats to transportation." The SPOT program is one means by which the TSA fulfills that responsibility, which enables the BDOs in the field to assess individuals exhibiting behaviors indicative of terrorist activity and referring them for additional screening or law enforcement officer (LEO) intervention. Many legal precedents and lawsuit outcomes are based on the interpretation of this document, most notably, the decision that participating in commercial transportation is not a constitutional right, and therefore, TSA screening is not a violation of privacy or civil rights.

3. Best Practices for TSA Behavior Detection

During the creation of SPOT, the TSA consulted with the FBI behavioral analysis unit, DHS behavioral sciences division, and subject matter experts (SME) in the field including Israeli security and El Al airlines. Published opinions on the subject range from "we should just do what Israel does" to more realistic opinions that recognize the scalability issues and unique challenges faced in the

[44] Security Administration, *Aviation and Transportation Security Act* (Washington, DC: Department of Homeland Security, Transportation 2001), http://www.tsa.gov/assets/pdf/Aviation_and_Transportation_Security_Act_ATSA_Public_Law_107_1771.pdf.

United States.[45] With numerous persons or entities claiming to be "experts" in the field, opinions vary on what the TSA's problems are; however, most sources agree that the interview process or using behavior detection techniques is a valid security technique, particularly because automated detection technology has a finite capability and is not infallible; therefore, it is also essential to search for intent, and not just "things."[46] Sources related to the subject of best practices include several credible experts from university professors to high-ranking Israeli security officials. Included in best practices is information from SPOT contributor Detective Lieutenant Peter J. DiDomenica at the Massachusetts State Police, which includes principles from the disciplines of physiology, psychology, neuroscience, specific research regarding suicide bombers, as well as a pioneer practitioner's perspective.[47]

4. Summary

The subject of deception detection (not behavior detection) is well documented on both sides and remains highly controversial, with test conditions and laboratory limitations proving to be a point of contention. Opinions vary widely on the use of such techniques in an airport environment. Dissenters cite privacy concerns, scientific cultural variance, fatigue, and concept of operations; while supporters cite lab results and anecdotal evidence from the law enforcement realm. The congressional pressure to prove the effectiveness of the program or lose funding will force TSA to improve technique, and eliminate any appearance of profiling. Both the negative and positive literature has provided valuable insight into the direction for a thesis on this topic.

---

45 *New York Times*, "Aviation Security and the Israeli Model," September 30, 2009, http://roomfordebate.blogs.nytimes.com/2009/12/30/aviation-security-and-the-israeli-model/?_r=0#more-27215.

46 *New York Times*, "What's Missing in Airport Security?" December 28, 2012, http://roomfordebate.blogs.nytimes.com/2009/12/28/whats-missing-in-airport-security/.

47 Ibid., 4.

C. RESEARCH DESIGN

This research is focused on improving the TSA's SPOT program to maintain a countermeasure that provides threat agnostic detection capability, as well as a valuable deterrence. The research includes identifying the strengths and weaknesses of the current SPOT policy and procedures, as well as an analysis of the recent GAO and OIG SPOT audits.

The traveling public is reliant on the TSA's multilayered, risk-based approach to aviation security. SPOT is one of the most important and unique layers available to the TSA in that it is threat agnostic, designed to detect intent versus things. The need to detect intent becomes more evident when considering the number of "things" that could threaten aviation security, and the technology and process required to detect those things. SPOT also provides an important element of deterrence, as no blueprint, published capability, or known state of the art is readily available, as with other automated physical screening equipment. Losing this layer to lack of funding or lack of understanding could leave a significant gap in the layered system, which leaves only the documented, finite detection capabilities of physical screening equipment.

The science of behavior detection is controversial at best. Although the paper addresses the basic principles and capabilities, as well as the documented criticism, it does not attempt to quantitatively prove the effectiveness of behavior detection. The science in general has both staunch proponents and equally staunch detractors; however, it is difficult to assess the effectiveness of the science accurately due to limited sample size and the inability to imitate the emotions associated with the fear, stress, or anxiety effectively that would accompany an attempt to commit a terrorist attack. The study addresses the GAO analysis, literature review, and conclusions regarding the TSA's SPOT program.

The focus of the paper is on TSA's implementation of behavior detection through the SPOT program, and whether the TSA can measure effectiveness

and implement strategic changes to add security value to the program. The GAO and OIG have independently made recommendations for SPOT, some of which are obvious and operational in nature and are not addressed at length. For example, both the GAO and OIG found inconsistencies with the TSA's data input systems and methods, which is an operational issue easily addressed and not within the scope of this study. Rather, this study attempts to determine if evidence is available to support the continuation of SPOT, as well as determining whether strategic improvements could add security value and credibility to this controversial program.

The scientific evidence available related to deception detection comes from published works by psychologists and behaviorists, academic journals from experts in the field, meta-analyses for test and evaluation of deception detection, and expert opinion from practitioners. This study attempts to ensure all results are understood within the context and conditions of the studies performed, rather than base conclusions on purely quantitative output data. Additionally, this study investigates the possibility of quantifying the deterrence value of this security measure, which has been mentioned by the TSA administrator.

Much of the summative data for the program comes from published, unclassified versions of GAO and OIG reports (provided by the TSA and assumed to be accurate). Also important to the study are other official government publications, such as sections of public law, Congressional reports, and expert Congressional testimony. Internal TSA news stories, as well as public published versions, are also used when necessary as qualitative support of program effectiveness or value added.

Two analyses were performed to reach the conclusions and recommendations of this study. The first was an analysis of the GAO and OIG reports that have drawn conclusions based on their own analysis. This portion is essentially an analysis of an analysis to determine if conclusions drawn by the GAO and OIG audits were based on robust, accurate analysis of the program data. The rest of the analysis was conducted using standard policy analysis

techniques to assess the strengths and weaknesses of TSA's SPOT program based on publicly available information. As the landscape of threat and risk mature, it is more important than ever for security policies and programs to evolve and improve to keep up with the changing environment. The use of behavior detection by the TSA has been in place since 2007; however, few changes have been made to the implementation or the strategic direction of the program. To determine if a new strategic direction can improve the SPOT program, a multi-goal policy analysis was conducted while implementing the existing analysis of the GAO and OIG throughout. Existing information was gathered and synthesized to form a baseline understanding of the current landscape. To provide scope and direction to the project, impact categories were derived from the literature, which include: 1) candidate selection, 2) training methods, 3) utilization concepts, and 4) evaluation mechanisms.

The study contains quantitative data used by both the TSA and GAO to assess the SPOT program effectiveness, such as annual budget, number of officers, interactions, referrals, and arrests, cost per referral, cost per arrest, etc. Much of the study is based on quantitative analysis derived from the literature, which explains why a certain method or recommendation would be beneficial. The study also focuses on missing data that may strengthen or alter some of the existing conclusions about SPOT. Additionally, sub-analyses are weaved into the recommendations that may address best practices of other agencies, and academic documentation that supports a recommendation. Qualitative analysis includes topics, such as why one training method is better than another, why meta-cognitive testing can improve performance, why certain human factors data may help improve concept of operations (CONOPS), and why deterrence is a valuable performance indicator, although maybe not measureable in the traditional sense. The project resulted in a defensible and contextually accurate assessment of the existing documentation, as well as actionable recommendations for strategic improvement to SPOT. The recommendations demonstrate a true understanding of the current landscape, provide a realistic

path forward while acknowledging that they are major strategic shifts, and require an entirely new framework for SPOT in the future. They are intended to be actionable recommendations for consideration and implementation by TSA/SPOT leadership and program officials.

D. OVERVIEW OF REMAINING CHAPTERS

The study begins with a brief background of SPOT, to include the history of the program, a look at the current state, and the current debate over the effectiveness of such techniques. Chapter III begins the assessment of the GAO and OIG conclusions from their respective audits. Chapter IV follows with a policy analysis reviewing the strengths and weaknesses of SPOT based on GAO and OIG audits, as well as publicly available information. Finally, in Chapter V, recommendations for SPOT are based on an analysis of the existing data and the available literature, with every attempt made to take the literature's conclusions in correct context and make the relevant applications to SPOT.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

To assess the SPOT program accurately, it is necessary to understand the background, not only operationally, but politically as well. This section also discusses the current status of SPOT.

A. HISTORY

The TSA has been practicing behavior detection techniques since 2004 when it launched behavior analysis experiments at Portland International Jetport and T. F. Green International Airport.[48] This program was a product of TSA collaboration with the Massachusetts State Police Department, which was operating a similar program called behavior assessment screening system (BASS). The BASS (and subsequently SPOT) program is primarily based on the research of Detective Lieutenant Peter J. DiDomenica, which includes principles from the disciplines of physiology, psychology, and neuroscience, as well as specific research regarding suicide bombers.[49] The basic premise is that a engaged in high consequence deception is experiencing fear (of discovery), stress, anxiety, or excitement leading up to and during the activity. This fear or anxiety present visually to a trained observer, as involuntary physical and physiological activity, such as increased heart rate, facial displays of emotion, changes in speed and direction of movement, nervous sweating, as well as other similar manifestations.[50]

The program also considers certain characteristics of physical appearance that indicate something is abnormal (for example baggy winter clothing in summer); these characteristics are not related to race, ethnicity, or religion. If the assessment of appearance and behavior lead the officer to believe something is

---

[48] *The TSA Spot Program: A Law Enforcement Perspective: Hearing before the U.S. House of Representatives Committee on Science, Space, and Technology Subcommittee on Investigations and Oversight*, 5.

[49] Ibid., 4.

[50] Ibid.

outside the norm, then the officer may also approach the subject with casual conversation, and then analyzes the subject's responses for anomalies (for example, level of comfort/knowledge about their circumstances, travel plans, etc.). Lieutenant DiDomenica describes the program using the acronym A-B-C-D, which stands for analysis of baseline, addition of a catalyst, and scan for deviations.[51] In the TSA world, a catalyst could be the screening process, a casual conversation with a BDO, the presence of a canine team, or anything that could potentially expose the subject's intentions (in the mind of the subject). The TSA simply describes the program as "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."[52] The "baseline" principle is of key significance in the practice of behavior detection. The baseline principle is simply knowing the everyday, common circumstances so well that anything outside of those parameters stands out. A good example would be an experienced bank teller easily picking out a counterfeit bill while counting a stack of cash. They are so experienced with the feel of real cash (the baseline) that they can instantaneously identify a counterfeit bill while rapidly peeling through a large stack of bills. Similarly, the practice of behavior detection relies on the theory that a well-trained and experienced officer conducting observations of thousands of "regular" individuals in an airport environment could easily identify a person who looked or acted in a manner outside the norm or baseline (i.e., threatening or malintent). In conducting observations, the officers are presumably familiar with the normal actions or reactions of both experienced and inexperienced travelers that comprise their standard population, and are aware that traveling in general, or specific traveling circumstances, may be stressful for some of the population.

---

[51] *The TSA Spot Program: A Law Enforcement Perspective: Hearing before the U.S. House of Representatives Committee on Science, Space, and Technology Subcommittee on Investigations and Oversight*, 4.

[52] United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 1.

The collaborative pilot programs eventually became the TSA SPOT program, and in 2007, the TSA created a new position called BDO designed to screen passengers at federalized civil aviation airports using BASS-like techniques.[53] The purpose of the program is to identify or deter "terrorists attempting to exploit TSA's focus on prohibited items and other potential security weaknesses."[54] This program could also be described as threat agnostic, as the TSA is screening passengers for intent versus screening them for prohibited items. Since the initial pilot, the SPOT program has developed its own unique CONOPS based on a point system. Point values are pre-assigned to specified behaviors, and then the BDOs cumulatively score individuals being evaluated. Their response to the individual (ignore, escalate to conversation, escalate to additional screening, or escalate to law enforcement) corresponds to different pre-determined point thresholds. The point system is unique to the TSA's application of behavior analysis and provides a framework that encourages objective application of the process while discouraging profiling or bias based on race, religion, appearance, etc.[55]

The GAO noted that the TSA deployed the SPOT program nationwide without "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."[56] With a significant budget (primarily for staff) and lack of empirical evidence that the program is effective, it is not surprising that the program has been the subject of major scrutiny. Over the past four years, the TSA's SPOT program has undergone two GAO audits and an OIG audit, which provides a total of 17 recommended actions to improve the

---

[53] United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 1.

[54] Ibid.

[55] Ibid., 10.

[56] Ibid., 14.

program. The respective recommendations are listed as follows, with the more strategically relevant (and addressed in this study) in bold (summarized):[57]

- Conduct a comprehensive validation study of the principles of SPOT.
- Conduct a comprehensive airport risk assessment to inform deployment of SPOT.
- Conduct a cost-benefit analysis of SPOT.
- Revise and implement strategic plan to include risk assessment (above) and cost (above).
- Study the feasibility of using checkpoint surveillance video of terrorists to understand behaviors.
- Improve data input procedures for the transportation information sharing system.
- Standardize BDO communication guidance with the transportation security operations center (TSOC).
- Direct TSOC to use all available law enforcement and intelligence databases when running LEO referral names.
- Establish a method to measure the effectiveness of the program and evaluate BDO performance.
- Establish data input controls to ensure complete, valid, accurate data is entered into the system.
- Systematically conduct evaluations of the SPOT training program.

The OIG audit acknowledged the GAO findings and added the following (summarized).[58]

- Develop a comprehensive strategic plan for SPOT.

---

[57] United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 60–62.

[58] Office of Inspector General, *Transportation Security Administration's Screening of Passengers by Observation Techniques (Redacted)*, 2.

- Develop and implement controls to ensure completeness, accuracy, AUTHORIZATION, and validity of referral data entered into the system.
- Develop and implement a plan that provides recurrent training for the BDOs and BDO instructors.
- Develop a plan to assess the BDO instructor's performance on a regular basis.
- Monitor and track the use of the BDOs for non-SPOT related duties.
- Develop a process for identifying and addressing issues that affect the success of the SPOT program, such as the selection, allocation, and performance of the BDOs.

B. CURRENT STATE

The SPOT program is organized into two offices, the threat assessment capabilities (TAC) division, which drives the program's strategy and direction, and the real time threat assessment division (TAD), which runs the day-to-day operations. SPOT has expanded to ~160 airports and now has more than 3,000 BDOs; additionally, the TSA has requested a Fiscal Year (FY) 2013 budget of $227 million, a 15 percent increase over five years.[59] As far as the FY15 budget is concerned, the TSA currently owes the GAO a response as to whether the SPOT program is effective at identifying terrorist intent. As such, Congress has not only reduced the FY15 budget, but also withheld $25 million of the approved budget pending an evidentiary response from the TSA that the program's behavioral indicators are effective. The budget hostage situation is potentially due to the TSA not providing a robust challenge to the GAO report findings; had they done so, it may be that accepting some of the recommendations while providing robust academic defense of the program may have ended the debate. The SPOT program seems to be suffering from an identity crisis, in part due to the GAO audit and related conclusions, but also from a lack of innovation and

[59] United States Government Accountability Office, *Transportation Security Administration: Progress and Challenges Faced in Strengthening Three Security Programs* (GAO-12-541T) (Washington, DC: U.S. Government Accountability Office, 2012), 6.

proactivity by the SPOT program. The GAO used academic literature regarding human lie detection capabilities to draw conclusions about SPOT. While the TSA did rebut the findings, they did little to differentiate SPOT techniques from human lie detection, and subsequently, the press and outside organizations now equate BDOs applying SPOT to human lie detectors.

Recently, however, the DHS did perform an independent evaluation of SPOT effectiveness, which yielded highly favorable results, but it was not delivered until well past the due date, and rejected by the GAO due to poor test construction. As it stands, it appears likely that unless the TSA can make substantive, strategic changes to the SPOT program, they may lose funding for it altogether.

1. Concept of Operations

The TSA operates SPOT in several configurations, with the most prevalent being simple observation and engagement of passengers in the security screening queue, and most recently, managed inclusion (MI) supporting TSA Pre™ operations at the airport. The observation of passengers includes the BDOs working in pairs, apart from one another but in communication, and observing large queues of passengers entering the checkpoint. In this case, the BDOs may engage passengers in casual conversation, or more directed conversation if they feel the passenger is exhibiting any signs of deception or indicating behaviors inconsistent with an aviation traveler. In the MI configuration, the BDOs work in conjunction with K9 teams or explosives trace detection (ETD) sampling teams to include (through lack of exclusion) additional passengers into the expedited TSA Pre™ screening lane. The BDOs also operate in a highly mobile configuration as partners in a visible intermodal prevention and response (VIPR) team. The team can consist of TSOs, BDOs and federal air marshals (FAM) working with state and local law enforcement to augment security forces in the transportation domain. These mobile teams are intended to be deployed to transportation sectors other than aviation, including rail, mass transit, pipeline,

etc.[60] The protocol in each of these configurations is essentially the same, and is more accurately described as the three locations or circumstances in which SPOT is conducted. The actual SPOT activity consists of the BDOs observing passengers as they approach different "stressors" of the screening process, such as a K9 team, an ETD machine, or a travel document checker (TDC). Faced with these stressors, passengers may begin exhibiting uncontrollable physiological behaviors (associated with fear of discovery), such as sweating, nervous behavior, or other behaviors inconsistent with the normal behavior of an aviation passenger in those same circumstances. These physical layers of security may illicit deceptive behavior from an approaching terrorist or criminal, and cue the BDOs to engage the passenger with either casual or directed questioning (according to their training), which may confirm or dispel the suspicious behavior. For example, passengers who exhibit suspicious behavior as they approach the TDC may be carrying a fake ID, or they may be nervous because they do not speak English. If observed by the BDOs, the BDOs may ask directed questioning that can reveal criminal activity, such as the possession of a fraudulent document, or can simply reveal that the passenger is a non-English speaker who is nervous about the potential communication issues at the TDC. The BDOs use their training and critical thinking to assess each situation using the totality of circumstances and conditions. Persistent or unresolved suspicious matters are directed to local law enforcement for further assessment.

Law enforcement agencies and anti-terrorist organizations worldwide, including the FBI, CIA, Bureau of Alcohol Tobacco Firearms and Explosives (ATF), state/local law enforcement agencies, and the "gold standard" of aviation security in Israel, use similar principles as an alternative to purely automated countermeasures. However, when the politically charged topic of domestic aviation security is added to the mix, it is necessary to provide meaningful justification for the program rather than "it works because it works." It is

60 "Visible Intermodal Prevention and Response (VIPR)," August 23, 2013, http://www.tsa.gov/about-tsa/visible-intermodal-prevention-and-response-vipr.

reasonable for lawmakers and budget czars to request program justification from the TSA; however, it is a complex task to prove the effectiveness of SPOT with such limited data available, and such a unique threat environment in which no attacks on domestic aviation have been conducted. Drawing conclusions about effectiveness based on the number of terrorists captured will have implications for most anti-terrorism organizations within the DHS. The GAO has stated in its audit reports that 1) "available evidence does not support whether behavioral indicators, which are used in the…..TSA.... SPOT program, can be used to identify persons who may pose a risk to aviation security,"[61] and 2) the literature review does not validate the practice/science of behavior detection.[62] While this thesis does not attempt to prove the controversial science of human lie detection, the next few paragraphs are dedicated to analyzing the GAO conclusions that led to these claims to determine whether the TSA SPOT program adds security value and should continue to operate in an airport environment.

[61] United States Government Accountability Office, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection*, 1.

[62] Ibid., 16.

III. ANALYSIS OF GAO CLAIMS

While the GAO and OIG have made several valid tactical recommendations, such as improving data collection, improving communication, and studying the viability of the use of close captioned television (CCTV), they were all precluded by the claim that SPOT techniques are not effective and that no evidence or literature exists to the contrary. Their analysis of the TSA SPOT data resulted in the conclusion that the TSA could not demonstrate the effectiveness of the program. However, additional research into the techniques indicates that they can work, so it must be determined if the TSA can take corrective action to make the practice viable in an airport environment. Most law enforcement agencies teach and use similar principles, and the DHS has an entire department dedicated to behavioral science. To determine if the GAO's conclusions are based on robust evaluation of the data, this analysis reviews several of their claims for context and validity. That is not to imply that the GAO has intentionally misrepresented any facts, or that the documentation it relied on is inaccurate, but rather to determine if contextual nuances exist within the data that could indicate alternative conclusions. While the GAO strives to be an independent, data driven organization, this analysis will determine the accuracy of their conclusions by focusing on the context of the data used.

A. SPOT HAS NOT DEMONSTRATED EFFECTIVENESS

A combination of claims that SPOT needs to 1) establish a method to measure performance of the BDOs, and 2) establish data input controls to ensure the accuracy of data, which leads this study to conclude that SPOT has a strategic metric issue. Collecting data with sufficient accuracy can be easily addressed operationally, but must be done before any data can be used and analyzed to indicate effectiveness. However, the larger problem is that SPOT needs to establish metrics that, when analyzed, would be indicative of performance. The lack of fidelity into SPOT performance has led the GAO to

withhold budget from SPOT, but worse than that, suggests that the TSA may not actually know how effective the program is. The TSA can demonstrate that SPOT regularly identifies criminals at the checkpoint, those intending to commit or in the act of committing a crime, such as smuggling, kidnapping, human trafficking, child pornography, etc. (see Figure 3). It is indicative of some success since persons engaged in high-stakes deception will react predictably to the fear of discovery. Unfortunately, it is not possible to obtain a true P(d) (using criminal identification as proxy data) because it is not possible to also know how many persons actively engaged in criminal or deceptive behavior have passed through the system without being caught. It is impossible to prove the negative, which makes it impossible to obtain a true P(d), and complicates the analysis for both the GAO and TSA. Suggestions on how to understand and begin measuring effectiveness are included in Chapter V.

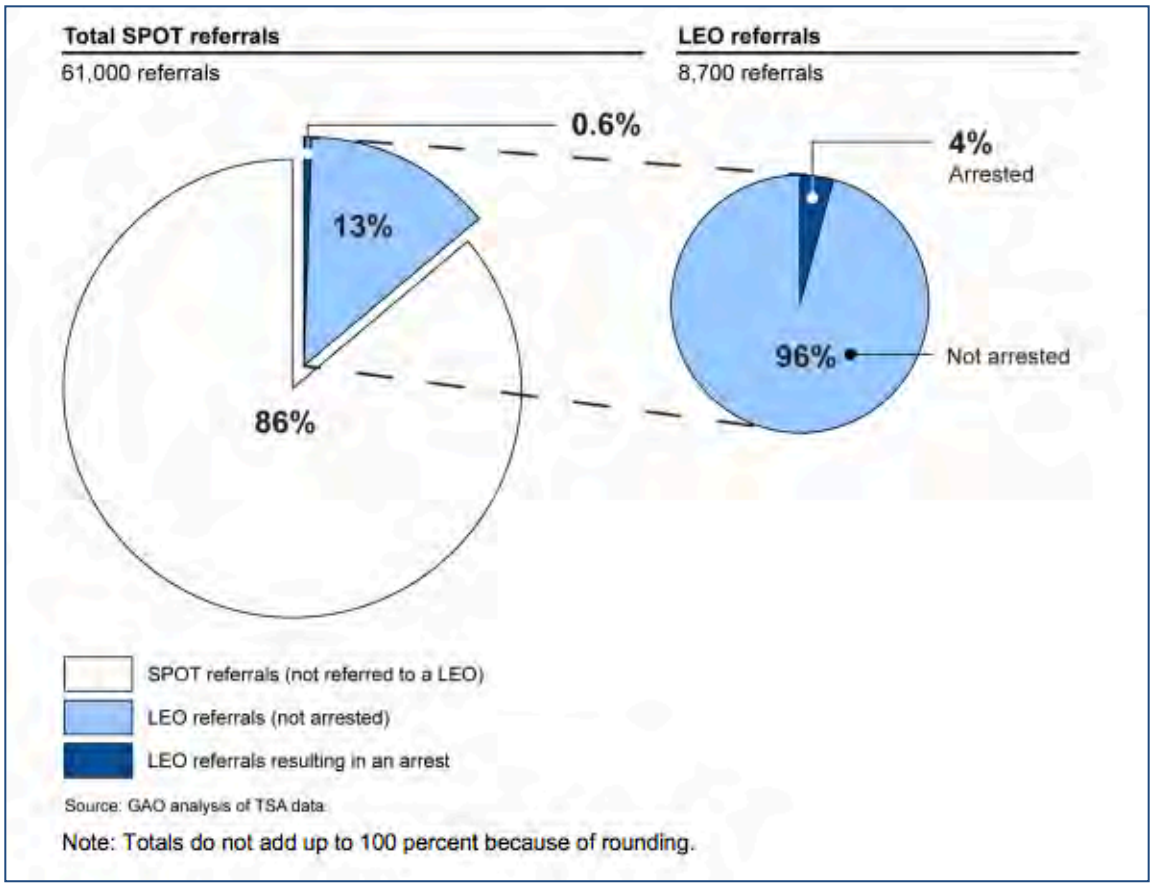

Figure 3. TSA SPOT Referral/Arrest Rates[63]

B. SPOT HAS NOT CAUGHT A SINGLE TERRORIST

The GAO notes that the SPOT program has not caught a single terrorist, which indicates that the program is not an effective anti-terrorist tactic. As a simple bullet point, the claim is certainly accurate, but taken in context may not be indicative of a lack of effectiveness. No known plots or attack attempts against domestic aviation have occurred since 9/11, which means no relevant data is available from which to draw conclusions about the effectiveness of any aviation countermeasure or program. Additionally, all TSA security programs (and most other domestic anti-terrorism programs) are operating in the same environment. Given the low frequency of terrorist attacks originating in the United States,

63 United States Government Accountability Office, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection*, 8.

success cannot be measured by the number of terrorists captured. By this logic, the United States should also stop inspecting luggage and persons at the airport and borders, since no terrorists have been captured using these methods. The logic used by the GAO is akin to saying that a designated hitter in a baseball game is ineffective because he got zero hits in the game. It may be true that he did not get a single hit, but that data point is not sufficient to draw conclusions about his performance. To assess his effectiveness, it is also essential to know how many at bats the hitter attempted. In both cases (SPOT and the baseball player) the data alone does not provide sufficient backing from which to draw conclusions about effectiveness. The low frequency of terrorist attacks (or lack of data) makes it difficult for either the GAO or TSA (really all of the DHS) to understand the effectiveness of anti-terrorist programs. However, it may be that SPOT and other security programs are providing an intangible level of deterrence, which makes aviation an unattractive terrorism target due to the adversaries' reduced likelihood of success, whether real or perceived.

Neither the TSA nor GAO knows how many terrorist plots against aviation have existed, been attempted, or possibly even been thwarted or deterred by the mere existence of the TSA. What is known is that zero successful terrorist attacks against domestic aviation has occurred, and while the DHS cannot necessarily attribute that success to any program (for the same reasons), it is certainly not appropriate to use it as evidence that a program does not work. If success were measured by the number of terrorists apprehended, then most anti-terrorist organizations would be considered ineffective. The lack of data does not correlate to a lack of effectiveness; however, it does indicate that proxy methods need to be developed to understand effectiveness.

C. SPOT FAILED TO INTERCEPT TERRORISTS ON 23 OCCASIONS

The GAO analysis also noted that 16 known or suspected terrorists (KST) have moved through eight SPOT airports no fewer than 23 times.[64] As a data point, it would seemingly prove that SPOT is ineffective; however, taken in context is not an indicator of performance. Several of the KSTs were leaving the country, and none of them was engaged in an imminent plot against domestic aviation, and as such, was not engaged in any deceptive behavior while traversing the airports. The principles of SPOT are based on the "fear of discovery," and a KST simply travelling from place to place would not be expected to exhibit behaviors consistent with deception. The initial statement of this point seems like reason enough to eliminate SPOT; however, taken in context, this data point does not support the implication that SPOT should have stopped these travelers.

D. SPOT IS APPLIED INCONSISTENTLY

The GAO also criticizes the program for having variable referral rates, noting that the secondary screening referral rates range from zero to 26 referrals per BDO per 160 hours worked, and LEO referrals range from zero to eight per 160 hours worked.[65] The TSA provided the data for analysis (Figure 4), which is presumably accurate; however, plausible reasons may exist for the variance in referral rates other than blatant inconsistency. To have high confidence in this claim, based on the data provided, the GAO may have been better served to consider the data at an airport level, checkpoint level, and even shift level. A particular airport may have more criminal activity occurring within their traveling population than another airport, which yields more referrals per BDO than another airport. Likewise, a particular checkpoint may have more criminal activity

---

64 United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 46.

65 United States Government Accountability Office, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection*, 6.

due to the arrival destinations served by the airlines at that checkpoint. For example, if one checkpoint has flights that leave for a destination known for drug activity, then a higher percentage of criminal activity might be expected in that checkpoint, which results in more referrals per 160 hours, as compared to a checkpoint servicing a low crime area. At the shift level, it could be that a certain international flight leaves only after 8 p.m. and contains a high percentage of selectees. Thus, the BDOs working those hours may expect to refer a higher percentage of passengers, even compared to the same airport and checkpoint during the morning hours. To draw conclusions about the variance in referral percentage, the GAO would want to also include data from the travelling populations of different airports, checkpoints, flights, etc., to determine if the variances were consistent with what could be expected from that traveling population. The fact may very well be that the variance in referral rates is due to inconsistency, but other relevant conditional data need to be considered when drawing that conclusion.

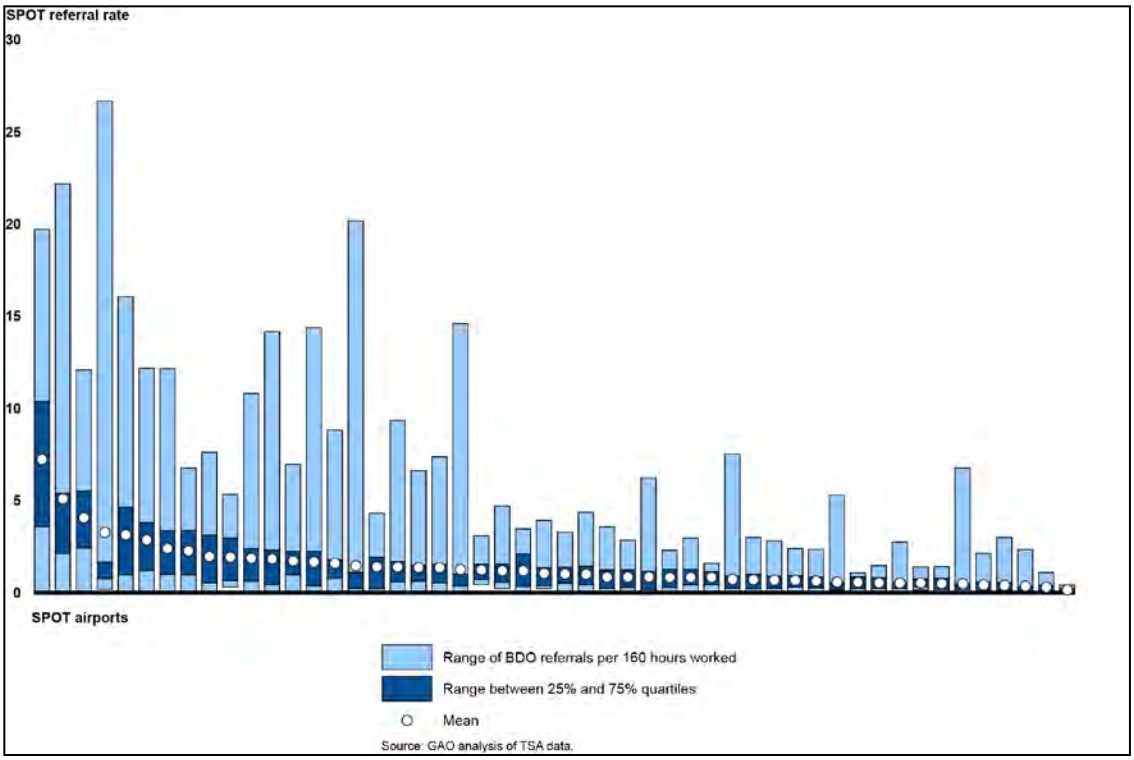

Figure 4. TSA Referral Rates among 49 Airports[66]

A more valuable data point to understand consistency would have been the arrest per referral ratio. The GAO noted (in a footnote) that this ratio ranged from zero to 17 percent of referrals amongst the 49 airports, but did not draw any specific conclusions from the data point. While several factors influence this outcome, such as local police policies on arrests for certain offenses, this data point does indicate a performance difference amongst the BDOs at different airports. This data is valuable as a comparison with other airports and even the individual BDO population. If all BDOs were performing at the same level of proficiency, then much smaller variance in arrest ratio or prohibited item ratio per referral (not provided by TSA) would be expected. For comparison purposes, no conclusions can be drawn about the ideal arrest or prohibited item detection per referral ratio, but it can be assumed that the higher the ratio, the higher the

66 United States Government Accountability Office, *Report to Congressional Requesters, Aviation Security: TSA Should Limit Future Funding for Behavior Detection Activities* (GAO-14-159) (Washington, DC: U.S. Government Accountability Office, 2013), 26.

accuracy of the BDO, with the total number of referrals being irrelevant. The GAO did not specifically call this out, but the variance in arrest to referral ratio is indicative of variable performance amongst BDO programs at the airport level. Further investigation could parse out individual BDO performance related to the arrest to referral ratio. The benefit of looking at performance or consistency from this angle is that the performance is measured by the outcome rather than the output, with the number of referrals being a meaningless output, and the ratio of arrests or prohibited item detection to referrals being a valuable outcome metric.

The GAO and OIG both made recommendations regarding the training and training instructors for SPOT, which suggests that the TSA provide recurrent training to instructors and consistently evaluate the instructor's performance. Using the variable referral data, arrest to referral ratio, instructor performance data, and the number of instructors and training locations together may have strengthened the conclusion that the TSA SPOT is applied inconsistently throughout the country.

E. LITERATURE DOES NOT VALIDATE THE PRACTICE

In 2008, the National Research Council of the National Academy of Sciences issued a report confirming that behavior and appearance monitoring might be able to play a useful role in counterterrorism. They also concluded that no consensus exists within the scientific community that these techniques are ready for use in the counterterrorism environment.[67] The state of research on the topic is thus summarized, with most studies, both for and against, concluding, "more research is needed." However, close examination of the academic research used by the GAO shows that it is using the performance of only "human lie detectors" as a proxy for SPOT performance, which is an apples-to-oranges comparison, and the studies compiled do not approximate the techniques used

[67] United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 15.

by SPOT. They are similar in nature, like apples and oranges are both categorized as fruit, but not sufficient to draw even basic comparisons.

The GAO conclusions rely primarily on two well-known meta-analyses as its basis of scientific evidence for SPOT. Meta-analyses, although they are great time savers, can be misused and misinterpreted. For example, the more studies they include, the less likely it is that the conditions and populations were sufficiently replicated across studies. Thus, the compilation is often less relevant than an individual study with contextually accurate test conditions. Some common academic criticisms of meta-analysis include the following.[68]

- One number cannot summarize a research field.
- They are easily biased by not including all studies in the field.
- Commonly mix apples and oranges (large variance in test conditions).
- Important studies are ignored.
- Replication of conditions is unlikely in multiple studies.

Critics of meta-analysis use the treatment of acute versus chronic pain symptoms as an example of how meta-analysis can result in misleading results. For example, if a treatment works great for acute symptoms, but is ineffective against chronic symptoms, then the combined meta-data would show the treatment is moderately effective. The "moderate" conclusion is inaccurate for both populations and conditions; the treatment is extremely effective in acute conditions, and completely ineffective in chronic conditions.[69] It is a classic example of a poor study selection. The Cleveland Clinic Journal of Medicine makes the point simply with the statement, "The outcome of a meta-analysis depends on the studies included."[70] A certain naiveté exists in drawing general

---

[68] Michael Borenstein, "Ch. 43 Criticisms of Meta-Analysis," in *Introduction to Meta-analysis*, ed. Borenstein et al. (Chichester, U.K.: John Wiley & Sons, 2009), 377–387.

[69] Ibid., 380.

[70] Estaban Walker, "Meta-analysis: Its Strengths and Limitations," *Cleveland Clinic Journal of Medicine* 756 (June 2008): 432, doi: 10.3949/ccjm.75.6.431.

conclusions about SPOT from a specific, bounded, limited set of studies. That is not to say that the meta-analysis does not serve a purpose, or that the developer of the meta-analyses intended to mislead anyone, but understanding the limitations of the meta-analysis may help the GAO add context to its conclusions.

For example, one of the primary meta-analyses used by the GAO is completely focused on studies containing interview techniques and the ability of a person to detect a lie during an interview with the conclusion that test subjects were little better than chance at detecting the deception.[71] SPOT is not a human lie detection program, and numerous issues have been raised with using this meta-analysis to draw conclusions about SPOT. For one, SPOT is not strictly an interview technique; in fact, the interview is a last resort, as the bulk of the practice is based on the observation of a person's appearance and behaviors, and the analysis of those cues against the context of the current circumstances. The BDOs have the advantage of being able to observe the passengers without necessarily being observed themselves, in which case the passengers are not inclined to hide any particular behavior or pattern. In the referenced meta-analysis, neither the subject matter, nor the interviewer, nor the interviewee correlates to those in a SPOT scenario. The BDOs uses their training and critical thinking to determine the appropriate resolution using the totality of circumstances; they do not interview every nervous passenger. During the GAO's two-year sampling period involving 49 airports, 365 of the 8,700 SPOT LEO referrals were arrested for fraudulent credentials, possession of weapons, warrants, etc.[72] While the ratio is not particularly impressive, it still represents hundreds of criminals removed from the transportation system. The untold portion of the story (in part due to poor data collection by the TSA) is that many interactions also end up with the confiscation of dangerous prohibited items, although the passengers may not be arrested for whatever reason. An

[71] United States Government Accountability Office, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection*, 51.

[72] Ibid., 44.

independent assessment of SPOT concluded that SPOT was nine times more effective at identifying persons engaged in criminal activity or possessing a dangerous item than the random selection of passengers for additional scrutiny;[73] however, the GAO has rejected these results citing a number of limitations. While the TSA acknowledges the limitations of the study, it stands by the macro level data, which indicates SPOT is substantially better at identifying high-risk individuals than random selection.[74]

Both the TSA and GAO face a dilemma when relying on results from lab testing, as most researches agree that lab research uses non-representative samples and low-stakes, trivial lies; the results of such tests are controversial, if not questionable. The liars used in the studies are mostly psychology students telling lies of no consequence (selected by the experimenter). The research participant pool designated as liars is not representative of the travelling public, and had no emotional stake in the lie or the outcome of the interaction with the lie catcher. The lie catchers frequently had no training whatsoever, and are not representative of trained SPOT practitioners or other professionals using these techniques.

DePaulo et al. note that cues to deception are stronger when lies are about transgressions,[75] which indicates that subjects forced to tell a lie of no consequence may not exhibit the same cues to deception as someone with a stake in the lie. In other words, smuggling a bomb through airport security will be more stressful (greater consequence) than smuggling a candy bar out of the corner drug store; thus, the high-stakes actor is more likely to exhibit irrepressible physiological responses and deceptive behavior. The studies did not attempt to replicate the heightened fear of discovery associated with committing a criminal

---

[73] *Behavioral Science and Security: Evaluating TSA's SPOT (Screening of Passengers by Observational Techniques) Program: U.S. House of Representatives, Committee on Science and Technology Subcommittee on Investigations and Oversight*.

[74] Ibid., 89.

[75] Depaulo et al., "Cues to Deception," 74.

or terrorist act, and therefore, the results of the studies appear to be limited in their applicability to SPOT.

Also of note, the meta-analyses do not refer to SPOT and were not compiled for the purpose of assessing SPOT. Re-creating the conditions of a real liar versus lie catcher scenario has plagued researchers for decades; however, newer research is attempting to focus on high-stakes lies by studying criminals and law enforcement officers in real scenarios.

The GAO analysis would have benefited from the inclusion of some high-stakes studies, such as O'Sullivan's 2009 meta-analysis compiling data from 31 police departments in eight countries, and concluding that high-stakes lies were detected at rate of 67.15 percent, while the low stakes lies were only detected at a rate of 55.17 percent.[76] Too many inconsistencies and dissimilarities within the GAO used meta-analyses exist to make any responsible comparisons with, or judgments of SPOT, not to mention additional studies and meta-analyses with contradictory findings.

F. THE IMPORTANCE OF CONTEXT IN DATA ANALYSIS (EVIDENCE)

When reviewing recommendations from the GAO audit of SPOT, it seems as though important contextual nuances were not considered when drawing conclusions from the data analysis. Many analyses, particularly difficult ones like behavior detection, can become overly focused on quantitative results or "outputs" (how many terrorists, how many arrests, how many false alarms, how many data input errors, etc.) in lieu of valid methods for assessing program effectiveness. While outputs can be reliable indicators of performance, they can also be a distraction in a security environment. Security practitioners must be more concerned with outcome than output. Outcome is difficult to measure. Thus, in many cases, organizations default back to things they can easily measure to claim success or failure. The GAO appeared to be strictly focused on the outputs of TSA SPOT provided data, and frequently draw conclusions without

[76] O'Sullivan et al., "Police Lie Detection Accuracy: The Effect of Lie Scenario," 533.

the proper context, and never focus on outcome. Based strictly on outcome, it could be argued that the TSA "system" is very successful not because it screens a lot of people and confiscates a lot of pointy objects, but because zero terrorist attacks have occurred on domestic aviation since 9/11. That is not to say the system is perfect, but it has certainly made aviation a less attractive target to an existing or would-be terrorist. The key element, which is hard to measure—and even harder to take credit for—is deterrence. It may be outside the norm for a government agency to claim deterrence as success, but it must be factored in assessing effectiveness. The bureaucracy should recognize the outcome of the system for the last 12 years, and begin to account for deterrence in a security agency's performance goals. In the current environment, if the TSA had caught two terrorists and missed one, a measureable 66 percent success rate would have occurred, which on paper would be pretty successful, and maybe even praised. So which is better? A 66 percent rate (not a bad output), or raising the barriers of a successful attack so high that terrorists are 100 percent deterred from targeting aviation? In the current state of bureaucracy, the latter receives the criticism of "having no evidence to support the effectiveness of the program," even though it meets the desired outcome. The TSA and GAO must resolve their cyclical disagreements by beginning to define success, and subsequently, measure it using a combination of outcomes and outputs.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. STRENGTHS AND WEAKNESSES OF BEHAVIOR DETECTION

Calling attention to the limitations of the meta-analysis used by the GAO is not intended to conclude that TSA SPOT needs no improvement and should continue as is. Many areas both strategic and operational require improvement to make the program more efficient and effective. In fact, the program has not undergone any major improvements since its inception in 2007, and the GAO and OIG made several logical operational recommendations that would indeed improve the program. However, most of those recommendations do not offer any strategic guidance or insight into how to achieve an outcome that would justify the existence of the program (a positive outcome would presumably be to improve the effectiveness and credibility of the program). Exploring what the program does well and poorly may reveal some strategic improvements that will help justify the program's existence, and more importantly, improve the chances of finding or identifying a terrorist.

A. STRENGTHS

1. Removing Criminals from the Transportation System

During FY2011 and FY2012, at the 49 airports studied by the GAO, TSA SPOT referred 61,000 passengers to secondary screening. Of the 61,000 referrals, 8,700 were further referred to law enforcement. Of the 8,700, 364 were actually arrested by local or federal law enforcement (see Figure 3). However, when dealing with statistics, it is important to understand the conditions and context of the numbers. For example, the arrest percentage is not very impressive if an arrest is the only valuable outcome. Moreover, an arrest is only one "successful" outcome, and one that TSA does not control. For example, a passenger referred to secondary screening may have been in possession of a prohibited item that the TSA or law enforcement confiscated. Confiscating dangerous or illegal prohibited items, such as knives, guns, or incendiary devices

(not out-of-policy water bottles or a small pair of scissors), is a positive outcome of SPOT for which no output measure exists; not only does keeping these items off airplanes makes everyone safer, but the discovery of the item is a direct result of SPOT intervention. Additionally, these confiscations (even of dangerous items) rarely lead to arrest, which means much of the effectiveness data from SPOT was not even available for consideration. Additionally, other reasons may explain why a passenger with a dangerous item may not be arrested, including the lack of law enforcement personnel to respond, and differing laws and regulations in different cities/states/jurisdictions. Arrests should be used as a proxy for measuring effectiveness, given the lack of terrorist attempts against domestic aviation, but should not be the only outcome measure for the effectiveness of the program.

Regardless of a person's point of view, SPOT kept 365 criminals (from the airports and time period under review) from boarding an aircraft, and countless dangerous items from getting on board. Criminal referrals and confiscated items may have value as a proxy for the ability to detect a terrorist (effectiveness) in the absence of actual terrorist attacks or attack data. In fact, the BDOs have also been known to discover the TSA "covert" testers intending to carry simulated IEDs through the checkpoint.[77] Subject to the similar "fear of discovery" (although to a lesser degree in the absence of high-stakes consequences) as a terrorist, the testers exhibit behaviors that are, at times, recognized by the BDOs.

2. Flexibility

Being a human-based capability, the BDOs have several advantages over technology-based capability. A key strength of the SPOT program is the ability to deploy anywhere, and to alter mission and mission location rapidly without significant startup, reposition, or shutdown time. The BDOs are not limited to locations with power outlets and lockable storage. The BDOs can and are

[77] Blogger Bob, "The Truth Behind the Title: Behavior Detection Officer," *The TSA Blog*, February 29, 2008, http://blog.tsa.gov/2008/02/truth-behind-title-behavior-detection.html.

frequently deployed at aviation security checkpoints, airport perimeters, air carrier gates, other areas of the airport, and even to rail locations courtesy of VIPR team participation. Due to the flexible nature of behavior detection, it is a key component of the TSA's risk-based security strategy. For example, randomly pulling out passengers for additional screening theoretically improves the P(d) of the system, but focusing that same level of effort on passengers behaving suspiciously makes even more sense than random selection. BDOs can also be deployed purposefully. For example, a flight may have two known high-risk passengers onboard, which could make it a higher-risk flight than normal; the BDOs could be easily deployed to the gate of said flight to observe all passengers boarding as an additional informed layer of security. Due to their flexible nature, the BDOs also bring about an opportunity for unpredictability. Being rapidly deployed to gates, or security identification display area (SIDA) access points, or other non-standard locations provides a layer difficult to plan for, increases overall security effectiveness, and adds deterrence value. For example, a BDO working the public area (prior to the checkpoint) of Orlando International airport identified a man exhibiting suspicious behavior. His identification of the man's behavior led to additional scrutiny, including searching his checked bags that resulted in the discovery of several unnamed suspicious items and the man's arrest by the FBI.[78] No other TSA layer of security is frequently active in the public area, and the flexibility of the program allowed this man to be identified well before the checkpoint and before he was able to cause harm to other passengers.

3. Deterrence Value

Deterrence has been mentioned frequently throughout this document, due to the immeasurable value that it provides. While it is true that TSA's SPOT program has not apprehended a single terrorist, a single terrorist attack aimed at domestic aviation has also not occurred. The outcome is that no attack attempts

[78] Blogger Bob, "Behavior Detection Officers Lead to Arrest in Orlando," *The TSA Blog*, April 04, 2008, http://blog.tsa.gov/2008/04/behavior-detection-officers-lead-to.html.

originating in the United States have taken placed, most likely due to the deterrence value of the layered security system in place. So does the number of terrorists apprehended dictate the success of a program? It may if performance assessment is inappropriately based on output measurements. However, the fact that no terrorist attacks have occurred indicates in large part that the system is working. That is not to say that the lack of attacks might not be the result of other reasons. Intelligence gathering, the strengthening of passport controlled entry, and the weakening of centralized terrorist networks plays a role as well. However, the fact remains that it is more difficult today to attack aviation security than it was 12 years ago. Advances in technology play a role, but programs like SPOT are essential to disrupting the planning of an adversary, and making aviation an unattractive target. Unfortunately for the TSA, it is difficult or impossible to determine how much of a role deterrence plays in the lack of terrorist attacks; however, it should be noted that raising the barrier to entry (for an attack) makes aviation a less attractive target. Using the lack of terrorist apprehension to indicate the performance of a security program would be like removing police patrols in low crime areas. Removing the police officers reduces the level of effort required to commit a crime, which suddenly makes the low crime area an attractive target for criminals. In a 2013 testimony to the House of Representatives, TSA Administrator John Pistole explained to Representative Mark Sanford (R-S.C.),

> There's no perfect science, there's no perfect art of this…This has been over seven years and we have screened by observation over 4 billion passengers, it actually comes out to 50 cents and in some instances 25 cents per passenger. To which the Congressman replied, In reverse, you could say, a billion dollars [spent] with no results." Mr. Pistole's final response: "I would say there's a result in terms of deterrence.[79]

[79] Ashley Halsey, "House Member Questions $900 Million TSA 'SPOT' Screening Program," *Washington Post*, November 14, 2014, http://www.washingtonpost.com/local/trafficandcommuting/house-member-questions-900-million-tsa-spot-screening-program/2013/11/14/ad194cfe-4d5c-11e3-be6b-d3d28122e6d4_story.html.

The Congressman's statement is indicative of applying the hermeneutic of a budget analyst, versus the more appropriate security context of outcome-oriented decision making. The Congressman is leaving out the context of understanding how many terrorist attempts have been made on domestic aviation security. A security system will catch zero terrorists if zero terrorist attempts are made. Mr. Pistole's statement is clearly focused on the outcome of having zero terrorist attacks, understanding the contribution of deterrence to security, as well as the SPOT's contribution to deterrence.

4. Threat Agnosticism

One of the key features of SPOT is the fact that the officers are trained to detect intent based on direct observation of the potential adversary's behavior. In other words, find the bomber not the bomb. Technology provides only a finite, inflexible detection capability. For example, an ETD) machine is designed to detect numerous types of explosive material, but cannot encompass all types of explosive material. Therefore, it is only useful if the adversary is using one of the materials that the technology is capable of detecting. Similarly, the walk through metal detector (WTMD) is very effective against metallic threats, but it stands to reason that it would not be successful against non-metallic explosive material carried on the body. The point is that each of these devices depends on a finite detection capability provided per government specifications and limited by the state of the art of technology (i.e., what capabilities exist). Therefore, the public is dependent on the probability that the adversary will choose threat types exactly what the TSA's machines are designed to detect. Technology has made this nation safer and further development should be pursued, but technology alone is insufficient. Equally important is maintaining a capability not dependent on threat type, but on recognizing a potential human threat. Additionally, behavior detection is difficult to counter, as a product specification is not available online or from the manufacturer, or for sale on eBay. The fact that so many misperceptions exist only makes it more perplexing to the adversary, and thus increases its deterrence value. In other words, it may be that an adversary could

download the user's guide (or buy the whole system) for a WTMD, ETD, and X-ray to inform an IED concealment method better, but it will not account for the threat agnostic practice of behavior detection. Adversaries will have a difficult time planning for the BDO when they do not know when, where, or how the countermeasure is being applied. Despite the congressional scrutiny of the U.S. program, the international community understands the value of a threat agnostic capability, as evidenced by the Israeli dependence on behavior detection techniques, and the existence of similar programs in several other countries. Simply put, SPOT offers the TSA the capability to detect any type of high-stakes threat, including nuclear, explosive, and biological, etc., by observing the expected behaviors of a high-stakes adversary going through the stress of the inspection process and the fear of discovery. No single detection technology can detect such a variety of threats. In fact, behavior detection by security officials and passengers identified Richard Reid (the shoe bomber) as a potential threat due to his purchase of a ticket with cash, disheveled appearance, blank or empty stare, luggage not consistent with travel plan, etc. These "red flags" garnered Reid additional scrutiny at the departing airport and he was not allowed to board the plane. Unfortunately, the scrutiny did not turn up the IED hidden in his shoes, but the point is that a type of behavior detection was successful in identifying his intent; finding the IED would have been the responsibility of the physical search officers.[80]

5. Unintended Benefits

BDOs, as trained observers, will always notice things that are outside the norm. In fact, the longer BDO practice their craft, the more baseline behavior data they possess, which they use to make more precise judgments in more diverse scenarios. This experience is valuable for more than just screening passengers. In fact, keen BDO observation and awareness have led to the discovery of many signs of passenger distress, including responding to

[80] Mary Sisson, "Richard Reid (British Militant)," *Encyclopedia Britannica Online*, March 5, 2014, http://www.britannica.com/EBchecked/topic/1975161/Richard-Reid.

passenger health issues, missing children, and even identifying victims of human trafficking. In July 2012, two BDOs in Miami identified a woman who was exhibiting unusual behavior. They approached her with casual conversation and she indicated that nothing was wrong. After further observation, they approached her again with some more specific questions that revealed she was being kidnapped. The BDOs' training and experience led them not only to identify the behaviors outside the norm, but also to continue to follow up until a satisfactory resolution was reached. In this instance, the BDOs likely saved the woman's life; her four captors were arrested on a variety of charges including kidnapping and unlawful detainment.[81] Such events are clear indicators that the totality of observation techniques are effective at identifying anomalous behavior. In fact, the anomalous behavior is sufficient for success; it is not necessary to determine the exact intent of the passenger during the interaction. Identifying anomalous behavior will prompt the BDOs to either perform additional screening, or refer the passenger to a LEO, which allows the other layers of security to confirm and resolve the anomalous behavior.

B. WEAKNESSES

Isaac Yeffet, former head of security for Israeli state airline El Al, and proponent of behavior detection, claims that the TSA's implementation of behavior detection is "worthless," and cites candidate selection and training as the reasons it does not work for the United States.[82] The GAO and OIG have provided Congress no fewer than 17 recommendations to improve the SPOT program. As previously mentioned, some of these recommendations are tactical in nature, but very necessary and insightful. Since improving data collection and communications do not require graduate level analysis, this analysis attempts to

81 Willard Shepard, "TSA Agents Rescue Kidnapped Woman," *NBC 6 South Florida*. July 31, 2012, http://www.nbcmiami.com/news/local/Kidnapped-Woman-Was-Rescued-By-TSA-Agents-at-MIA-Authorities-164345996.html.

82 Dennis Schaal, "TSA's Behavioral Detection Skills Are 'Worthless,' Says Israeli Aviation Security Expert," Skift, November 13, 2013, http://skift.com/2013/11/15/tsas-behavioral-detection-techniques-are-worthless-says-former-el-al-security-director/.

focus on weaknesses that have strategic impact to the program or program effectiveness. Some of the weaknesses included in the following sections were expressly, or more often, generally discussed by aviation experts, as well as the GAO and OIG, and are noted accordingly.

1. Utilization

The media has joked at times that the TSA stands for Thousands standing around. This situation is quite possibly the result of the BDOs literally standing around. While they are actually conducting their observations, the appearance is that they are doing nothing. For both political purposes (the optics) and efficiency purposes, the TSA needs to ensure the BDOs are multitasking. The special training they receive should not preclude them from also contributing to the processing of passengers, or other additional duties, and may actually help them engage with more passengers more frequently. SPOT officers are currently engaged in three CONOPS, including SPOT operations at the checkpoint, VIPR teams, and MI, but each includes only the performance of basic SPOT procedures and the perceived change in CONOPS is really just a change in location. While valuable from a flexibility standpoint, it does nothing to improve the optic of them standing around or the efficiency at the checkpoint. Furthermore, the BDOs are hired into full-time positions, but the airport does not operate in eight-hour shifts and is only busy during peak times a few hours a day. The BDOs may only be “observing” (working) for a couple hours of their actual eight-hour shift, which would be fine if they were assigned BDO duties during peak times and other duties on non-peak times. Unfortunately, they are not performing other duties during their down time, which makes the program appear quite inefficient.

2. BDO Selection Criteria

Currently, the TSO pool is the only source of candidates for the BDO position.[83] No aptitude testing or psychometric testing is applied other than what is required to become a Transportation Security Officer (TSO). The BDOs must only meet the following hiring requirements (the requirements of a TSO) and are eligible to apply for BDO after one year as a TSO.

- have a high school diploma or general educational development (GED) credential **OR** at least one year of full-time work experience in the security industry, aviation screening, or as an X-ray technician
- be proficient in the English language (i.e., able to read, write, speak, and listen)
- be a U.S. Citizen or U.S. national at time of application submission
- be at least 18 years of age at time of application submission
- pass a drug screening and medical evaluation
- pass a background investigation including a credit and criminal check
- no default on $7,500 or more in delinquent debt (but for some bankruptcies)
- selective service registration required[84]

Although not attempting to validate the science behind behavior detection, the OIG study astutely notes that limiting BDO recruitment to current TSOs may not provide the program with the most qualified candidates.[85] Obviously, the aforementioned criteria do not in any way qualify a person to observe and assess behavioral and physiological indicators of a passenger (maybe the one year of experience in the airport environment is helpful).

---

83 Schaal, "TSA's Behavioral Detection Skills Are 'Worthless,' Says Israeli Aviation Security Expert," 11.

84 "Transportation Security Officer (TSO)," accessed October 7, 2014, https://www.usajobs.gov/GetJob/ViewDetails/383141100.

85 Ibid., 11.

More than a weakness, it should be more accurately described as a flaw in strategy of the program. To improve credibility and effectiveness, the TSA must address its weak hiring practices for the BDO population.

3. Training

The TSA currently requires that the BDOs undergo a four-day classroom training session, followed by three days of on-the-job training (OJT). During these sessions, BDO candidates must memorize all the program specific behaviors and appearances, along with all the associated point values, and pass a job knowledge test consisting of multiple choice, true/false, and case-based scenarios.[86] In May 2011, the TSA also began providing refresher training for existing BDOs in response to a job task analysis that indicated observation skills are perishable.[87] Both the GAO audit and the OIG investigation astutely recognized weaknesses within TSA's SPOT training program when they recommended that the TSA evaluate SPOT training periodically, offer refresher training to both BDOs and BDO instructors, and have a means to evaluate BDO instructors. While all three are logical and valuable, it seems as though they are recommendations to address symptoms of the true problems. The sources of these symptoms are 1) training inconsistency, and 2) inadequacy of training content.

a. Training Inconsistency

One of the most important aspects of specialized training for a large geographically dispersed population is consistency. At the time of the 2010 GAO audit, the TSA had 54 SPOT instructors scattered across the country. While it makes sense from an efficiency standpoint to have trainers at airports or regional locations, it can be detrimental to the primary outcome of producing BDOs who

---

[86] United States Government Accountability Office, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, 14.

[87] Office of Inspector General, *Transportation Security Administration's Screening of Passengers by Observation Techniques (Redacted)*, 9.

will operate consistently across all airports. In 2012, the TSA found that several of their BDO instructors "did not have the instructor knowledge, skills, or abilities to instruct BDO classes."[88] While the TSA did provide remedial instruction to those BDO instructors, the audit brought to light the inevitable variances present within a large pool of instructors, which results in inconsistent BDO performance in the field. Besides the lack of consistency (or difficulty in producing consistent results) and potential variance in quality, maintaining proficiency for 54 instructors is quite challenging and the infrastructure needed to refresh, evaluate, and maintain these instructors may be enough to eliminate any efficiencies gained. The potential value of the SPOT security layer and politically charged nature of this program demand that it be executed with precision, which requires nothing but the highest quality training standards.

b. Inadequacy of Training Content

Behavior detection is the most specialized function of TSA screening and likely requires the most cognition of any of the TSA screening functions. The TSA cannot expect a four-day training class to prepare a TSO adequately to observe a passenger's behavior, appearance, and demeanor, and quickly analyze the totality of the circumstances to determine if the passenger poses a threat to aviation security. Isaac Yeffet claims El Al Airlines trained their BDOs for lengthier periods of time (than the TSA) including three weeks on the job followed by rigorous testing.[89]

Many cultural, political, and socioeconomic variables may impact a BDOs' biases, and they need to be aware of the variables, as well as their own natural responses to them, versus their required professional response. In fact, BDOs, as well as all human beings, are susceptible to allowing their own biases to override or influence professional decisions. As BDOs have been criticized for

88 Office of Inspector General, *Transportation Security Administration's Screening of Passengers by Observation Techniques (Redacted)*, 9.

89 Ibid.

profiling in the past,[90] these biases and responses require thorough understanding; training in these areas can help ensure objectivity of passenger assessment.

The BDOs may also be confused about their authority or lack thereof. Given that the BDOs single out passengers exhibiting deceptive behavior who may actually be a threat, the BDOs need to be well versed in what they can and cannot do or say. Recently, a selectee passenger traveling from Minneapolis St. Paul International Airport (MSP) to Denver International Airport (DEN) was not given "enhanced" screening before departing MSP (although he did go through standard screening). He was met by a cadre of BDOs who were tasked with re-screening him at his destination. The passenger was understandably resistant to being screened AFTER his flight, and the BDOs appeared unsure how to handle his questions and refusal to be rescreened.[91] The BDO's response to the passenger's questions indicated that they were unsure of the authority (or in this case lack thereof, as the ATSA requires screening prior to boarding[92]) under which they were operating, and likely had not been trained on how to handle a situation effectively in which the passenger refuses to cooperate. The default response seems to be to call a supervisor, and the event revealed a gaping hole in the knowledge of BDO authorities.

4. Effectiveness and Performance Testing

One of the keys to success for any program is to be able to prove and track effectiveness. SPOT has difficulty with this task for several reasons that include not being able to measure deterrence or prove the negative (of how many deceptive passengers were missed if any), limited terrorist activity with

[90] Mark Hanrahan, "Logan Airport Racial Profiling Allegations: TSA Officers Complain Colleagues Are Targeting Minorities," *The Huffington Post*, August 12, 2012, http://www.huffingtonpost.com/2012/08/12/logan-airport-racial-profiling_n_1769648.html.

[91] Barry Donegan, "TSA Tries to Pat Down Man After His Flight, Watch Him Refuse," Ben Swann Truth In Media, September 12, 2014, http://benswann.com/tsa-tries-to-pat-down-man-after-his-flight-watch-him-refuse/.

[92] "Screening Passengers and Property," https://www.tsa.gov/sites/default/files/assets/pdf/49_usc_chapters_401_to_501.pdf.

which to verify effectiveness, unwillingness of critics to accept identifying criminals as a proxy for identifying terrorists, and limited methods to test effectiveness as true behavioral indicators cannot be replicated without the true fear of consequence. Other programs, such as technology programs, can have documented pass/fail testing protocols based on strict criteria or requirements set by the federal government prior to purchasing a piece of security equipment. A piece of technology generally has a target detection rate (of the specified subject matter) and a target false alarm rate to maintain a specified level of effectiveness without losing efficiency. These requirements are easy to test and can be "certified" in a lab environment by going through the checklist and ensuring all the technical requirements were met. Behavior detection cannot be tested with these traditional methods, as no detection requirements or false alarm requirements exist to measure against, and lab testing is unable to replicate the human emotion associated with high-stakes deception. This layer cannot be measured using traditional means, and lab testing is unable to replicate the type of human (subject) response that would be present in a real-world, high-stakes scenario, such as smuggling a bomb through the checkpoint.

THIS PAGE INTENTIONALLY LEFT BLANK

V. RECOMMENDATIONS

A. UTILIZATION RECOMMENDATIONS

- Recommendation—Utilization: Perform human factors evaluation of high performing interview officers to determine which psychometric attributes consistently appear. Open hiring to general public, using these attributes as core competencies for BDO hiring.

A key to the success of the TSA's controversial SPOT program is to ensure that the workforce is comprised of the best possible candidates, ideally with aptitude for the task. A 2008 study on deception detection noted that the average person is slightly better than random at identifying when a person is lying or trying to deceive.[93] However, the study also identified several groups that were much better at detecting deception than the average person. The groups included therapists, Secret Service agents, and other law enforcement types. The Secret Service agents were among the highest performing, presumably because of their training on scanning large crowds for non-verbal cues (similar to a BDO's function at an airport), and the low frequency of uncovering a plot or perpetrator (not predisposed to assuming a crime has been committed).[94] Although other types of law enforcement did well, the commonality of frequently dealing with criminals may predispose them to a higher instance of type 1 errors (false positives). The study also noted that certain individuals had a "genius-level" aptitude for detecting deception, which indicated the need to identify their common characteristics and use them to develop specific core competencies to be used as BDO qualifications.[95]

The TSA should conduct human factors aptitude research to include psychometric testing (using industry recognized methods, such as Myers-Briggs or similar) to determine which traits occur most frequently in highly skilled

[93] Frank, Menasco, and O'Sullivan, "Human Behavior and Deception Detection," 5, sec. 2.

[94] Ibid., 5–6, sec. 2.1.

[95] Ibid., 6, sec. 2.2.

behavior assessment officers (not limited to the TSA BDOs, but should include officers from other government agencies or industries using similar techniques). For example, a Meyers-Briggs assessment measures psychological preferences in the form of four dichotomies: extraversion/introversion, sensing/intuition, thinking/feeling, and judging/perceiving to come up with a personality type.[96] As an experiment, the TSA's human factors branch could select high performing BDOs (using some form of logical performance data as criteria, such as referral to arrest ratio) and administer the Meyers Briggs to determine what patterns are consistent in the psychological predispositions of high performing BDOs. If patterns emerge (such as strong marks toward sensing versus intuition, or thinking versus feeling), then the TSA can incorporate them into hiring criteria, or use it as a basis for further research or even custom aptitude test development. Additionally, a 2012 study in the *Journal of Forensic Psychiatry and Psychology* notes that research suggests certain characteristics, such as age, profession/experience, and handedness/hemispheric dominance, can play a role in deception detection aptitude.[97] Attempting to find the most appropriate candidate for a job is not a new concept. The Department of Defense uses their armed forces vocational aptitude battery (ASVAB) to determine the vocational aptitude for new military recruits.[98] While a basic aptitude test, the TSA should consider a similar model to find the best possible candidates for SPOT. Candidate selection is a key element for the effectiveness and consistency of the program moving forward, and finding candidates with appropriate aptitude should be a priority for the TSA. Hiring best qualified/prepared applicants based on psychometric assessment of aptitude and scientifically proven core competencies also provide much needed defensibility and credibility to the program, which is a major shift in strategic hiring practices for the TSA, and will

96 " MBTI® Basics," accessed October 8, 2014, http://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/.

97 Shaw, Porter, and Brinke, "Catching Liars: Training Mental Health and Legal Professionals to Detect High-stakes Lies," 2.

98 "ASVAB Test Explained," accessed November 15, 2014, http://www.military.com/join-armed-forces/asvab/asvab-test-explained.html.

likely be met with many hurdles, including union pushback. However, the TSA must be committed to solving the hard problems if it wants to maintain this layer of security.

- Recommendation—Utilization: Place additional hiring emphasis on candidates possessing program-enhancing characteristics, such as language skills and cultural competency/background.

Diversity many times plays only a political role in the workplace, in which employers check the box that they met their Equal Employment Opportunity (EEO) and affirmative action quotas. However, following the intent of diversity rather than the letter of diversity can actually improve the performance of the organization by moving them a step closer to cultural competency. Cultural competency refers more to the cognitive state of organizations, and how they view and react to the different cultures in their communities (keeping in mind culture is not limited to race, and also includes other elements, such as belief systems and other social identity markers). Cultural competency is one of the most important factors for success in this globalized society, whether it is to better expand the customer base to improve the bottom line, or for the fire service to know that Somalis maintain numerous generations in a single household. In either case, understanding the "community" of interest provides valuable information. Particularly in the security, health, law enforcement, fire services, and other community services, this competency can even save lives.

The Georgetown's National Center for Cultural Competence provides a conceptual framework for culturally competent organizations, which includes acquiring and institutionalizing cultural knowledge, and adapting to the diversity and cultural context of the communities they serve.[99] This same competency should exist within the BDO ranks according to the community needs and breakdown of the airport's traveling community. The traveling community is not always simply the areas surrounding the airport, but refers to the cultural

[99] Georgetown University for Child and Human Development, "Conceptual Frameworks/Models, Guiding Values and Principles," accessed November 15, 2014, http://nccc.georgetown.edu/foundations/frameworks.html#ccdefinition.

breakdown of the traveling population frequenting the airport. In the aviation security industry a popular expression is, “if you’ve seen one airport, you’ve seen one airport.” The statement embodies the idea expressed in this thesis, as each airport has its own set of unique circumstances including its specific trend of diversity from the traveling community, and the hiring selections should reflect the needs of the traveling community on a per airport basis. For example, airports in south Florida should have a high concentration of Latino BDOs, as well as a gender breakdown similar to the traveling community. The Latino BDO concentration is necessary not only because of the Latino demographic in south Florida, but also because many of the airports in south Florida fly to Latin American countries. In this case, both the local and the traveling communities have a high Latino concentration. While a simple example, each airport would need to consider its unique circumstances and build a BDO workforce according to the needs of its own traveling community. It is not necessary for every BDO to speak Spanish in Florida; that would be an extreme interpretation. The idea is that assets are available that have knowledge of additional languages, religions, ethnic background, etc., to build true cultural competency in the BDO ranks. Cultural competency is a buzzword typically used in a training context, in which organizations (particularly other public service organizations) pile diversity information on their existing workforce and expect them to become “competent” based on that training. While training in this area can be valuable, this study is suggesting that the TSA take that concept a step further by building a culturally competent baseline by making an effort to hire from within the communities that the TSA protects. Bureaucracy often prevents government leaders from making innovative decisions, and will require commitment from leadership to achieve the extensive modification to hiring practices.

- Recommendation—Utilization: Create “peak time” BDO position to ensure BDOs are available during the busiest airport traffic hours.

In the 2010 GAO audit, Dr. Paul Ekman (a major contributor to TSA’s development of SPOT) indicated that fatigue is a well-known issue amongst workers whose job requires intense observation and focus. He recommends

ongoing research to determine the duration of effective observation.[100] While an important recommendation that TSA should consider, fatigue could also be addressed through a strategic shift to hire part time BDOs.

The concept and practice of using peak time employees is frequently used in other industries, such as banking, where the bank needs more staff during peak times than they have full time equivalent (FTE). Rather than carry additional FTE, banks will hire peak time tellers to fulfill the need during the rush, but not have to pay additional staff to sit around during off peak hours (utilization problem).

Most airports operate in a peak time fashion during which the majority of their passengers leave on certain days and during certain hours. For example, Monday mornings and Friday afternoons typically see the highest volume of passenger due to business travelers commuting to and from work. Generally speaking, the majority of Monday morning passengers will arrive at the airport between 5:30 am and 8:30 am, which creates a large influx of passengers for about three hours. Hiring part-time BDOs for these peak shifts will allow the TSA to reduce their FTE BDO count and reduce the fatigue factor of BDOs by limiting the amount of time they must be engaged. Using BDOs when and where they will have the most impact also aligns with TSA's new risk-based approach to aviation security. Alternatively, to retain high performing BDOs who need full time employment, SPOT could be made a collateral duty utilized during peak times, with the officer conducting normal screening operations at off-peak times. In either case, the TSA should consider the five factors influencing travel patterns (day, time, season, holiday, and destination)[101] when determining how to best provide behavior detection at the most effective times while maintaining an efficient workforce. A peak-time or collateral work schedule that mirrors peak

[100] U.S. Government Accountability Office, Aviation Security, *Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*.

[101] Ed Hewitt, "A Peek at Peak Travel," IndependentTraveler.com, accessed October 8, 2014, http://www.independenttraveler.com/travel-tips/travelers-ed/a-peek-at-peak-travel.

airport times/seasons will provide the best combination of coverage and efficiency. Generally speaking, if normal SPOT coverage is 16 hours per day (assuming 1,500 BDOs per shift, and two shifts), and peak travel time is only six hours a day (assuming three hours in the AM and three hours in the PM), the SPOT program could cut man-hours by more than half. The assumptions in this exemplar are quite conservative, since in reality, peak hours are not consistent on a daily basis and only apply to the busiest travel days. In actuality, the TSA could consider a SPOT skeleton crew on non-peak days and seasons, to “right size” the SPOT FTE count, or reassign the BDOs to other duties during this time to minimize the SPOT hourly cost. Of course, it is a general discussion of the efficiencies to be gained and variances will occur among individual airports according to layout, checkpoint design, seasonality, etc., but the concept of peak time BDO will save thousands of man hours per day, and will allow the TSA to streamline the SPOT budget while still adding maximum security value. This practice would also improve the effectiveness of the program by limiting the duration of the task (with high cognitive load), which would presumably allow the BDOs to perform at a higher level during their limited shift. Additionally, the pool of applicants may be increased to include retired military or law enforcement officers who may already possess some aptitude but do not desire to work full time. Lastly, having the skillset of an “off-duty” SPOT officer imbedded in normal screening procedures during off peak times only improves the capabilities at the checkpoint. The peak time BDO concept also improves defensibility for the TSA, as it is able to allocate resources more efficiently and save taxpayer dollars by reducing dependency on FTE BDOs.

Of course, this concept is not without challenges, such as maintaining training or proficiency of a peak-time or part-time BDO. However, the seemingly erratic nature of air travel actually follows predictable trends, which allows the TSA to determine what a seasonal or peak-time schedule might look like. Once these patterns are established, the TSA can build in appropriate job performance

standards to include skill maintenance, refresher training, and continuing education for the peak-time/part-time population.

B. TRAINING RECOMMENDATIONS

- Recommendation—Training: Merge TSA developed BDO training with FLETC developed and administered training to provide a stronger baseline to the BDOs. Recommend FLETC behavioral science division review TSA training materials to determine whether the BDOs can simply attend an interview course, or if FLETC can modify interview courses to accommodate the BDOs.

According to TSA Deputy Administrator John Halinski, the TSA is in the process of centralizing most of their training into the Office of Training and Workforce Engagement (OTWE) to maintain and enhance the capabilities of TSA employees while providing a consistent training experience.[102] To accomplish this goal, in April 2012, the TSA established a training presence at DHS's FLETC in Glynco, Georgia. Although a component of the DHS, FLETC is an inter-agency training organization that has trained more than 1,000,000 law enforcement officers from 91 different government agencies or partner organizations.[103] FLETC's approach to instruction is to maintain "a mix of permanent, detailed, and recently retired staff (to) provide an appropriate balance of training expertise, recent operational experience, and fresh insight from the field."[104] Not surprisingly, FLETC maintains a behavioral science division responsible for designing and administering courses in the many aspects of human behavior. One of the principal topics it addresses is interview skills for criminal investigators and law enforcement officers. Of the topics covered in these classes, several would be directly applicable to the BDO position including interviewing strength and weakness forum, eye accessing cues and behavioral baselines, question

---

102 *Written testimony of Transportation Security Administration Deputy Administrator John Halinski for a House Committee on Homeland Security, Subcommittee on Transportation Security hearing titled "Eleven Years After 9/11 Can TSA Evolve To Meet the Next Terrorist Threat?,"* 112th Cong. (2012).

103 "Welcome to FLETC," accessed September 2, 2013. http://www.fletc.gov/.

104 Ibid.

types to elicit admissions, subject elimination interviews (ruling out a threat), and cognitive interviews. While it will require dedication to collaboration, it fits perfectly into the TSA training strategy of providing a consistent professional experience to all TSA employees. The TSA will need to consider whether the entire training can be done at FLETC, or if it will still need a small course focused on CONOPS to be taught locally. Either way, the TSA can address some of the GAO and OIG's valid complaints regarding underperforming instructors and inconsistent BDO performance across the country. Additionally, the TSA will be able to leverage existing instructional design and instructor expertise on the topic while potentially building a new course that may be desirable to other agencies or countries. This strategic shift has other intangible benefits as well, such as creating an esprit de corps and sense of camaraderie amongst the workforce. It is no coincidence that the members of the armed forces have all attended a "basic training" after which their sense of accomplishment bonds them together and creates a sense of pride. The TSA can improve the effectiveness of the program by providing a more consistent training experience administered by professional instructors and practitioners at a highly respected educational institution. Leveraging the expertise and reputation of FLETC's 40-year law enforcement training heritage also provides credibility and defensibility to the program, as FLETC's credentials have never been in question. Additionally, the TSA may be able to re-focus existing BDO instructors by assigning them the lower-skill tasks of providing local update training or facilitating continuing education as necessary. While a major strategic shift for the SPOT program, it coincides perfectly with the agencies new effort to centralize training for consistency, while leveraging existing government resources.

- Recommendation—Training: Add content to the basic BDO training to include understanding the authority under which they operate, cultural, political, and socioeconomic variables that affect a person's behavior; and how personal biases affect a BDO's response to those variables.

The inadequacy of training content mentioned in "weaknesses" refers to providing information about why the BDOs are required to do certain things. It is

strategic shift in training, as most TSA training is technical in nature and focuses on operations, such as what buttons to push and how to start up or shut down a system. The BDOs, however, are not pushing buttons and deciphering colors on an X-ray screen; their job is much more about psychology and understanding people and their normal behaviors in an airport environment. This type of skill cannot be taught with simple operational techniques, but rather needs to be taught from a perspective of understanding.

For example, recent events (the Denver incident mentioned earlier) have shown that the BDOs are often tasked to do things that might be outside their authority, or at least they may not know if they have the authority to do what they are doing. For instance, during the passenger re-screening incident in Denver, the passenger asked several times if he was being detained or arrested, and whether he was required by law to submit to the re-screening even though he had already reached his destination. The BDO in charge should have had clear answers to those questions, for himself as much as the passenger. If the BDO had been trained properly on what authority he was operating under, he would have known that chapter 449 of Title 49 C.F.R states, "…screening for domestic flights shall take place before boarding…" and that the TSA may not have had legal authority to rescreen that passenger after his flight had landed. Furthermore, the BDO was unable to discuss the concepts of legal detention versus arrest, neither of which was within his authority, which is a failure of training on the part of the SPOT program, and not specific to this individual BDO.

Understanding how cultural, political, and socioeconomic variables impact passengers' appearance, behavior, and response to questioning, should be a major focus of SPOT training. While the BDOs will learn through experience over time what the specific cultures and reactions are at their airport, a strong baseline in cultural competency would be beneficial. One way to identify the differences between people and their behaviors is to understand the similarities. As an example of improving training content, the BDOs curriculum could focus on basic physiological/psychological elements, such as the six universal emotions as

described by Dr. Paul Ekman. Dr. Ekman contends that all cultures respond similarly to the same basic human emotions, with the big six being anger, happiness, surprise, disgust, sadness, and fear (see Figure 5).[105] An understanding of what all cultures have in common provides a valuable baseline for performing BDO duties.

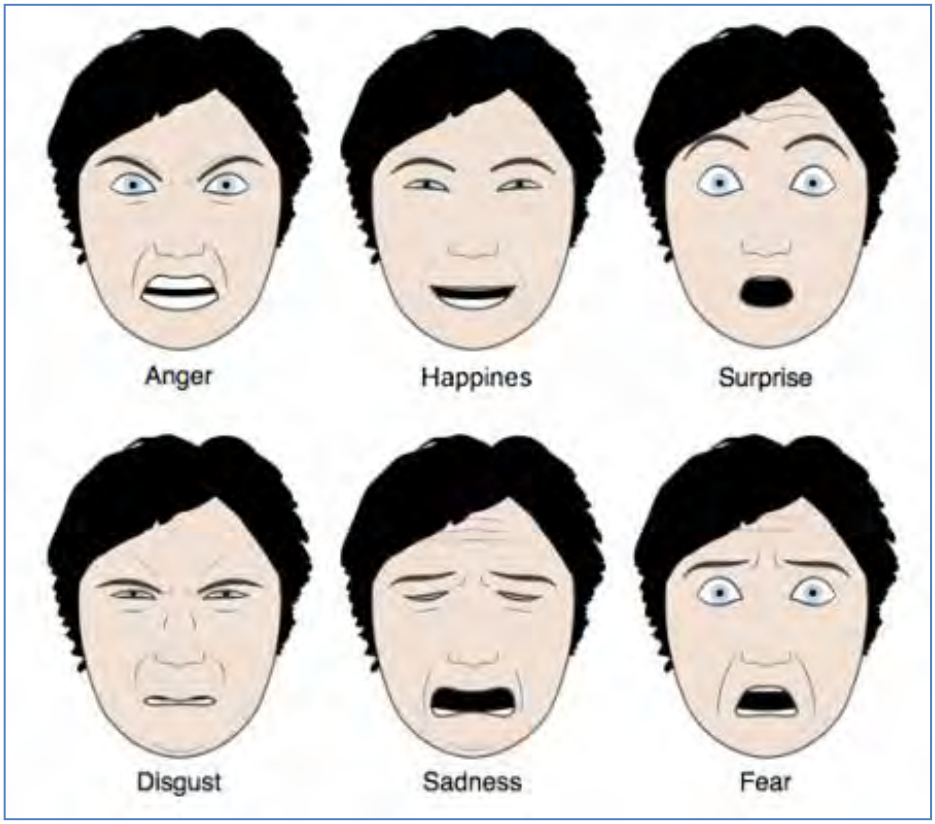

Figure 5. Six Universal Emotions[106]

Along with understanding universal emotions, additional discussion and content related to the cognitive biases that all humans are subject to, and how that may impact a BDO's work should happen. In particular, some understanding

105 Paul Ekman, "Are There Basic Emotions," *Psychological Review* 99, no. 3 (February 10, 1991), http://www.communicationcache.com/uploads/1/0/8/8/10887248/are_there_basic_emotions.pdf.

106 "Six Basic Emotions," June 26, 2013, https://managementmania.com/en/six-basic-emotions.

of confirmation bias, in-group bias, and observational selection bias would be beneficial. Confirmation bias generally means that people agree with people that agree with them, and in the case of a BDO, can lead to discriminatory reactions against misunderstood groups, which may then lead to profiling or the appearance of profiling.[107] Similarly, in-group bias (also called in-group favoritism), showing preference to a group sharing values, culture, interests, etc., and separating groups into "we and they" can affect BDOs' evaluation of individuals they do not identify with, which can lead to prejudice during their interaction.[108] Being aware of this and other types of group biases will help the BDO population make impartial decisions when dealing with a wide variety of "groups." Another bias to be aware of is observational selection. Sometimes an event like purchasing a new car can trigger a person's brain to start noticing the exact same car frequently. Observational selection bias makes it possible to believe that these cars have just begun appearing, when in reality, they have always been present, but had not been noticed prior to the purchase of the car. Similarly, a BDO who successfully identifies a criminal or terrorist may start to believe that their frequency of arrival is increasing, or worse yet, it may lead them also to impose out-group status to an entire religion, race, or group based on the discovery of one criminal or person with malintent. Cognitive bias is a major contributor to the failure of human based assessments, and the more aware the BDOs are of these biases and how to combat them, the less chance exists for profiling (even inadvertently) certain groups of people. This bias has to be balanced with data, as well. For example, a high percentage of Muslims being selected by the BDOs for additional screening at a Chicago airport does not necessarily indicate profiling as ~2.8 percent of the population in Illinois is

---

[107] Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2, no. 2 (December 18, 1997): 175, http://psy2.ucsd.edu/~mckenzie/nickersonConfirmationBias.pdf.

[108] Samuel L. Gaertner and John F. Dovidio, *Reducing Intergroup Bias: The Common Ingroup Identity Model* (Philadelphia, PA: Psychology Press, 2000), 5.

Muslim.[109] Further, the traveling population from Chicago airports could have an even higher Muslim population. Local TSA needs to be fully aware of the airport's traveling population and understand how the correlations between population breakdown and arrests, as well as other factors that influence which groups are selected for secondary screening most frequently, such as non-English speakers not understanding the restrictions at the airport.

Many cognitive biases occur, and the need for highly specialized, robust training is obvious when considering the legal issues, cognitive biases, and the variety of group variables faced by a BDO. As previously mentioned, this training is also best given to only persons with aptitude for this job requiring high cognitive function, the ability to minimize biases, and understand the local passenger population culture.

- Recommendation—Training: Offer advanced training classes in areas that will add value add at the checkpoint, as well as offer career advancement opportunities for the BDOs (by collecting and demonstrating proficiency in advanced training areas).

As previously mentioned, the BDO community is sometimes considered to be "standing around" and not assisting with passenger processing at the checkpoint. While technically just a perception issue, they should be capable of multitasking even if they are not assisting the checkpoint with processing passengers. A concept that would add value to the BDO program would be to have BDOs competent in advanced skills that could benefit not only the checkpoint but also the entire airport. For example, a BDO could elect or earn the opportunity to obtain first responder certification, cardiac pulmonary resuscitation (CPR) certification, "active shooter" certification, etc., to possibly even include a law enforcement status (of course, it would require reclassification).

This recommendation crosses over a couple of different weaknesses including training and utilization, but offers a strategic improvement to capability,

[109] Jahnabi Barooah, "Most and Least Muslim States in America (PHOTOS)," *The Huffington Post*, June 27, 2012, http://www.huffingtonpost.com/2012/06/27/most-and-least-muslim-states_n_1626144.html.

perception, and effectiveness for the program. On November 1, 2013, a lone gunman walked into terminal three at LAX and opened fire, and eventually killed one TSA officer and wounded two others.[110] In this instance, it would have been beneficial for TSA BDOs to have any of the advanced skills mentioned above. The recommendation is not suggesting that the BDO receive a 10-minute on-line course about CPR or active shooters, but be selected for advanced training in these areas based on performance/capability. While not only a training recommendation, it is also a CONOPS career progression recommendation, where being selected for, and obtaining any of these advanced training certificates will allow the BDOs to advance in their career or obtain additional pay for the additional skills obtained. Per previous recommendations, much of this information can be taught at FLETC using existing curriculum and courses. Much discussion of how the BDOs could have helped during the LAX shooting has occurred, including recommendations offered by the TSA and GAO during the November 14th hearing before the House of Representatives subcommittee on Transportation Security, and by the Los Angeles World Airport Police in their March 2014 after action report.[111] The recommendations include operational issues, such as improving communications, having more police on site, and using CCTV. All good recommendations, but none take advantage of an already present workforce (the BDOs) frequently putting themselves in harm's way by singling out suspicious individuals. Training these employees (assuming it is possible to ensure this workforce is the "right" workforce) is a great opportunity to enhance security at the checkpoint, as well as the value of the SPOT program. It is difficult to rely on emergency services or tactical response when they are

---

110 CBS/AP, "LAX Shooting Kills TSA Officer, Wounds Others," *CBSNews*, November 1, 2013, http://www.cbsnews.com/news/lax-shooting-kills-tsa-officer-wounds-others/.

111 *Written testimony of TSA Administrator John Pistole for a House Committee on Homeland Security, Subcommittee on Transportation Security hearing titled "TSA's SPOT Program and Initial Lessons From the LAX Shooting,"* 113th Cong. (2013) (testimony of John Pistole), http://www.gpo.gov/fdsys/pkg/CHRG-113hhrg87373/pdf/CHRG-113hhrg87373.pdf; Los Angeles World Airport Police, "Active Shooter Incident and Resulting Airport Disruption, A Review of Response Opeations," March 14, 2014, http://www.lawa.org/uploadedFiles/LAXLAWA%20T 3%20After%20Action%20Report%20March%2018%202014.pdf.

minutes away and hundreds of passenger's lives may be at stake at the checkpoint area. Having the capability to respond to medical emergencies (as first responders) and other critical incidents up to and including a law enforcement type response is a strategic shift that the TSA should consider for SPOT. That is not to say that SPOT needs to convert into a combination fire department and police department, but that those capabilities can be strategically placed at checkpoints for a small portion of the workforce. This placement serves to improve capability and credibility of SPOT, as well as incentivize the workforce to perform well by providing a career progression. The previously mentioned examples of the BDOs responding to medical emergencies and criminal activity demonstrate that the BDOs are positioned (literally and figuratively) perfectly for these additional responsibilities, as they are already observing the area for anomalous behavior or activity, and are frequently engaging the public.

C. EFFECTIVENESS AND PERFORMANCE TESTING:

- Recommendation—Performance: Collaborate with agencies training spy craft or undercover techniques to test the BDOs covertly or overtly.

One principle noted by Frank, Menasco, and O'Sullivan is that detecting the behavioral indicators is easier when the stakes or consequences are higher.[112] This principle could indicate that behavior detection in a security or criminal type environment will be more effective than in, for example, a therapist's office or academic study. It may also indicate that the SPOT program's effectiveness in identifying criminals and criminal activity does correlate to the effectiveness of identifying a terrorist who would be experiencing similar reactions to the fear of being caught in a high consequence scenario. Therefore, it is possible to test the effectiveness of the program (or a single BDO) if the TSA can provide controlled test subjects who truly do have a fear of being discovered due to an actual consequence. Other agencies, such as the CIA, ATF

[112] Frank, Menasco, and O'Sullivan, "Human Behavior and Deception Detection," 6, sec. 2.2.

and FBI, train their agents in espionage and undercover techniques including the art of deception. For this training to be effective, the tester needs to face a real consequence, incentive, or motivation, such as fear of dismissal or remediation. For this reason, it is recommended that the TSA work with these agencies during their agent's training phases, during which the trainee must avoid BDO detection to pass a class or move to the next phase of his or her training. This testing will not only provide a true test for a BDO, but also provide valuable feedback on the agencies' training candidate. This concept would improve effectiveness for SPOT by actually measuring effectiveness and identifying areas for improvement. The concept also improves defensibility for the program by providing actual performance data, which is also crucial to the cost-benefit analysis requested by both the GAO and OIG. This recommendation is based on inter-agency collaboration and resource leveraging to improve efficiency and provide a benefit for multiple agencies, which should be undertaken under the watch of human experts to ensure this type of incentivized testing will yield relative effectiveness data. Keep in mind that even these incentivized testers will not have their very life at stake, as would a suicide terrorist. This cross-pollination effort should address every aspect of SPOT including appearance, behavior, response to questioning, etc., so as not to not replicate an already flawed version of "interview testing," as described in the numerous deception detection meta-analyses. Ideally, if this method proves to be viable, it should also be capable of producing the number of negatives (missed adversaries), as well as provide insight into the most functional CONOPS for assessing the most passengers.

- Recommendation—Performance: Establish an operational baseline performance metric for existing BDOs using arrests/prohibited item to referral ratio.

While the agency collaboration and incentivized testing solution is likely a long-term and complex undertaking, the TSA could begin baseline performance for BDOs using existing metrics in a variety of ways. During the SPOT validation study, the DHS used the outcome to referral ratio to compare SPOT performance to random selection of passengers for interview, for which the outcome was the

combined number of arrests and dangerous prohibited items confiscated. The TSA can use this same methodology to compare the performance of all existing BDOs to start the process of establishing a baseline for BDO performance. Calculating the rate at which a BDO referral is found to be in possession of a serious prohibited item, or arrested by a LEO, is truly indicative of how well a BDO separates passengers into their respective risk categories. While existing data is not available to determine what percentage a BDO should be able to achieve, the comparative analysis of the BDOs will provide a range of TSA BDO capabilities, with a higher ratio indicating higher performance. Based on the capabilities achieved and the variance amongst the BDOs, the TSA will be able to establish a minimum requirement for this ratio and use that requirement as a baseline performance metric that the BDOs will need to achieve to maintain BDO status. This method will create new complications, but using a ratio allows the TSA to obtain relevant data from all BDOs regardless of their airport circumstances (i.e., high referral airports can still be compared to low referral airports). More investigation is needed to determine if operational, incentivized testing will also be useful; however, in the meantime, this metric can be established immediately, and followed by analysis to determine an appropriate range for the ratio. The value in this metric is also in the TSA holding itself accountable for performance, and being able to report that the BDOs must maintain “X” ratio of accuracy in deception judgments based on their referral to outcome ratio. This particular metric is low hanging fruit for the TSA, and analysis should begin immediately to find the baseline ratios for individual BDOs, as well as the program itself. The TSA cannot take this task lightly, as the follow on tasking, creating and holding the workforce to a minimum ratio, is a strategic shift and must result in the elimination of the BDOs unable to maintain an acceptable ratio. This task should be undertaken simultaneously with other recommendations, so that by the time a baseline is determined, it will be also be known what characteristics make a good BDO, and potentially, whether the improved content and consistency of training can improve the performance of

lower performing BDOs who may otherwise be eliminated. The TSA will need to sort the details of accountability, associated minimums outputs if any (such as a minimum number of referrals), probationary periods, remediation or mentorship as necessary, etc. Sorting these issues is trivial compared to the strategic and political value in understanding the effectiveness of SPOT, even if only in a relative context.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. CONCLUSION

The TSA SPOT program must continue to evolve, as it is one of the few layers that provide a real time threat assessment outside of the known technology countermeasures. Adversaries are constantly studying the TSA's countermeasures and are exploring mechanisms and concealment techniques to defeat them; however, it is far more difficult for terrorists to conceal their involuntary emotional, physiological, and physical responses that will occur while engaged in high-stakes deception. While the TSA should continue to pursue the basic corrective action needed to fulfill the GAO and OIG recommendations, it should also be considering the strategic direction of the SPOT program, including how to improve effectiveness and defensibility to make SPOT a legitimate, contributing, aviation security program. Considering creative solutions to complex problems aligns with the TSA's commitment to becoming a high performing organization.[113]

This most revealing part of this study is the inconclusive and controversial nature of the academic literature in this field. Essentially, no literature evaluating behavior detection in totality is available; the majority of the literature focuses on lie detection in an interview setting, or deception detecting with similar experiment conditions. Additionally, this study finds the academic literature in this field to be too dependent on studies using only trivial lies and unmotivated lie tellers as experiment conditions and test participants. Conclusions about SPOT should not be drawn from academic research relying on these conditions.

While this field of study collectively acknowledges the limitations of the experiment conditions, more research in the field of high-stakes lies and deception detection in an airport environment is needed to make conclusive claims about TSA SPOT. Areas in which the literature is in general agreement

[113] *Eleven Years After 9/11 Can TSA Evolve To Meet the Next Terrorist Threat?: Hearing before the U.S. House of Representatives Committee on Homeland Security, Subcommittee on Transportation Security*.

are: 1) verbal and non-verbal cues to deception do exist, 2) no “Pinocchio’s nose” telltale indicator of deception exists, 3) deception can be easier or more difficult to detect depending on the skill of the liar, 4) high-stakes lies may be easier to detect than trivial lies due to the powerful emotions associated with a motivated lie, 5) cues to deception may be more evident during personal lies, and 6) lie catchers can be trained to elicit indicators from liars by increasing their cognitive load.

This study concludes that SPOT is a valuable layer of aviation security, but needs to be evolving with the academic research and enhancing their capabilities as more is learned about the science. The GAO, OIG, and now this thesis, have analyzed SPOT with similar findings. The strategic shifts recommended in this thesis align with the TSA’s current efforts to become a more effective and efficient organization, and are intended to address the strategic gaps of the program. The TSA should consider major changes to SPOT to improve the security value and provide credibility and defensibility to this misunderstood program. A summary of the recommendations is provided as follows.

- Establish an operational baseline performance metric for existing BDOs using arrests/prohibited item to referral ratio.
- Develop hiring criteria based on a study of psychometric and other attributes of high performing BDOs.
- Place additional hiring emphasis on candidates possessing program-enhancing characteristics, such as language skills and cultural competency/background.
- Conduct all SPOT training at FLETC using established procedures.
- Revise SPOT curriculum to include explanation of TSA authority, cultural, political, and socioeconomic variables that affect a person’s behavior, and how personal biases affect response to those variables.
- Offer advanced training classes in areas that will add value at the checkpoint, as well as offer career advancement opportunities for

the BDOs (by collecting and demonstrating proficiency in advanced training areas).

- Collaborate with agencies training spy craft or undercover techniques to test BDOs covertly or overtly.

It is imperative that the TSA continue to develop the SPOT program to maintain a threat agnostic and unpredictable layer difficult for adversaries to "game." The SPOT program also provides a level of deterrence that would be lost if the program were defunded or minimized. Addressing strategic items, such as performance metrics, candidate selection, utilization, and training consistency, will not only appease political sensitivities, but also improve the security value of the program. While the TSA is making progress on some of the operational GAO recommendations, it should consider the recommendations in this thesis to provide direction and long-term stability for SPOT.

Additionally, areas for future research that may provide valuable strategic direction to TSA SPOT and other anti-terrorism countermeasures include:

- **Creating measurable metrics for deterrence**—Deterrence is a main goal of terrorism countermeasures but is not typically included in strategic plans as a measurable metric or outcome. Understanding deterrence may help organizations determine when to shift countermeasures, and when a countermeasure can be deescalated.

- **Developing alternative CONOPS**—TSA may want to explore alternative CONOPS to address some of the operational issues. For example, developing a cadre of LEO BDOs may alleviate the current LEO response issues, and may be a more viable concept for intentionally interacting with higher-risk individuals. Conversely, training everyone in behavior detection techniques may provide additional coverage in order to evaluate more passengers with these methods.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Barooah, Jahnabi. “Most and Least Muslim States in America (PHOTOS).” *The Huffington Post*, June 27, 2012. http://www.huffingtonpost.com/2012/06/27/most-and-least-muslim-states_n_1626144.html.

Blogger Bob. “Behavior Detection Officers Lead to Arrest in Orlando.” *The TSA Blog*, April 04, 2008. http://blog.tsa.gov/2008/04/behavior-detection-officers-lead-to.html.

———. “The Truth Behind the Title: Behavior Detection Officer.” *The TSA Blog*, February 29, 2008. http://blog.tsa.gov/2008/02/truth-behind-title-behavior-detection.html.

Borenstein, Michael. “Ch. 43 Criticisms of Meta-Analysis.” In *Introduction to Meta-analysis*. Edited by Michael Borenstein, Larry V. Hedges, Julian P. T. Higgins, and Hannah R. Rothstein, 377–387. Chichester, UK: John Wiley & Sons, 2009.

CBS/AP. “LAX Shooting Kills TSA Officer, Wounds Others.” *CBSNews*, November 1, 2013. http://www.cbsnews.com/news/lax-shooting-kills-tsa-officer-wounds-others/.

Depaulo, Bella M., James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. “Cues to Deception.” *Psychological Bulletin* 129, no. 1 (2003): 74–118. doi: 10.1037/0033-2909.129.1.74.

Donegan, Barry. “TSA Tries to Pat Down Man After His Flight, Watch Him Refuse.” Ben Swann Truth in Media, September 12, 2014. http://benswann.com/tsa-tries-to-pat-down-man-after-his-flight-watch-him-refuse/.

Ekman, Paul. “Are There Basic Emotions.” *Psychological Review* 99, no. 3 (February 10, 1991): 550–53. http://www.communicationcache.com/uploads/1/0/8/8/10887248/are_there_basic_emotions.pdf.

———.*Emotions Revealed: How Recognizing Faces and Feelings Improves Communication and Emotional Life*. New York: Times Books, 2003.

———. “Lie Catching and Microexpressions.” In *The Philosophy of Deception*, edited by Clancy W. Martin. Oxford: Oxford University Press, 2009.

Flora, Carlin. “When to Go With Your Gut.” *Psychology Today*, May 1, 2007, www.psychologytoday.com%2Fcollections%2F201205%2Fgo-your-gut%2Fyour-best-guide.

Frank, Mark G., Melissa A. Menasco, and Maureen O'Sullivan. "Human Behavior and Deception Detection." In *Handbook of Science and Technology for Homeland Security*, vol. 5. John Wiley and Sons, 2008. Accessed September 1, 2013. http://www.cl.cam.ac.uk/~rja14/shb10/frank2.pdf.

Gaertner, Samuel L., and John F. Dovidio. *Reducing Intergroup Bias: The Common Ingroup Identity Model*. Philadelphia, PA: Psychology Press, 2000.

Gladwell, Malcolm. *Blink: The Power of Thinking without Thinking*. New York, NY: Back Bay Books, 2007.

Greenfeld, Liah. "Are Human Emotions Universal?." *Psychology Today: Health, Help, Happiness + Find a Therapist*, April 23, 2013. http://www.psychologytoday.com/blog/the-modern-mind/201304/are-human-emotions-universal.

Grohol, John. "Psychology Secrets: Most Psychology Studies Are College Student Biased." *Psych Central.com*. Accessed November 9, 2014. http://psychcentral.com/blog/archives/2010/08/26/psychology-secrets-most-psychology-studies-are-college-student-biased/.

Halsey, Ashley. "House Member Questions $900 Million TSA 'SPOT' Screening Program." *Washington Post*, November 14, 2014. http://www.washingtonpost.com/local/trafficandcommuting/house-member-questions-900-million-tsa-spot-screening-program/2013/11/14/ad194cfe-4d5c-11e3-be6b-d3d28122e6d4_story.html.

Hanrahan, Mark. "Logan Airport Racial Profiling Allegations: TSA Officers Complain Colleagues Are Targeting Minorities." *The Huffington Post*, August 12, 2012. http://www.huffingtonpost.com/2012/08/12/logan-airport-racial-profiling_n_1769648.html.

Hartwig. Maria, and Charles F. Bond. "Why Do Lie-catchers Fail? A Lens Model Meta-analysis of Human Lie Judgments." *Psychological Bulletin* 137, no. 4 (2011): 643–59. doi: 10.1037/a0023589.

Hewitt, Ed. "A Peek at Peak Travel." IndependentTraveler.com. Accessed October 8, 2014. http://www.independenttraveler.com/travel-tips/travelers-ed/a-peek-at-peak-travel.

Los Angeles World Airport Police. "Active Shooter Incident and Resulting Airport Disruption, A Review of Response Opeations." March 14, 2014. http://www.lawa.org/uploadedFiles/LAX/LAWA%20T3%20After%20Action%20Report%20March%2018%202014.pdf.

ManagementMania.com. "Six Basic Emotions." June 26, 2013. https://managementmania.com/en/six-basic-emotions.

Matsumoto, David. “The Origin of Universal Human Emotions.” October 18 2014. http://davidmatsumoto.com/content/NG%20Spain%20Article_2_.pdf.

Military.com. “ASVAB Test Explained.” Accessed November 15, 2014. http://www.military.com/join-armed-forces/asvab/asvab-test-explained.html.

Myers & Briggs Foundation, The. “MBTI® Basics.” Accessed October 8, 2014. http://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/.

National Center for Cultural Competence. Georgetown University for Child and Human Development. “Conceptual Frameworks/Models, Guiding Values and Principles.” Accessed November 15, 2014. http://nccc.georgetown.edu/foundations/frameworks.html#ccdefinition.

New York Times. “Aviation Security and the Israeli Model.” September 30, 2009. http://roomfordebate.blogs.nytimes.com/2009/12/30/aviation-security-and-the-israeli-model/?_r=0#more-27215.

New York Times. “What’s Missing in Airport Security?” December 28, 2012. http://roomfordebate.blogs.nytimes.com/2009/12/28/whats-missing-in-airport-security/.

Nickerson, Raymond S. “Confirmation Bias: A Ubiquitous Phenomenon in Many Guises.” *Review of General Psychology* 2, no. 2 (December 18, 1997): 175–220. http://psy2.ucsd.edu/~mckenzie/nickersonConfirmationBias.pdf.

Office of Inspector General. *Transportation Security Administration’s Screening of Passengers by Observation Techniques (Redacted)* (OIG-13-91). Washington, DC: Department of Homeland Security, 2013.

Porter, Stephen, and Leanne Brinke. “The Truth About Lies: What Works in Detecting High-stakes Deception?.” *Legal and Criminological Psychology* 15, no. 1 (2010): 57–75. Accessed September 3, 2014, doi: 10.1348/135532509X433151.

Russell, James. “Is There Universal Recognition of Emotion From Facial Expressions? A Review of the Cross Cultural Studies.” *Psychological Bulletin* 115, no. 1 (1994): 102–41.

Schaal, Dennis. “TSA’s Behavioral Detection Skills Are ‘Worthless,’ Says Israeli Aviation Security Expert.” Skift, November 13, 2013. http://skift.com/2013/11/15/tsas-behavioral-detection-techniques-are-worthless-says-former-el-al-security-director/.

Security Administration, *Aviation and Transportation Security Act*. Washington, DC: Department of Homeland Security, Transportation 2001. http://www.tsa.gov/assets/pdf/Aviation_and_Transportation_Security_Act_ATSA_Public_Law_107_1771.pdf.

Shaw, Julia, Stephen Porter, and Leanne Ten Brinke. “Catching Liars: Training Mental Health and Legal Professionals to Detect High-stakes Lies.” *Journal of Forensic Psychiatry & Psychology* 24, no. 2 (January 17, 2013): 4, doi: 10.1080/14789949.2012.752025.

Shepard, Willard. “TSA Agents Rescue Kidnapped Woman.” *NBC 6 South Florida*. July 31, 2012. http://www.nbcmiami.com/news/local/Kidnapped-Woman-Was-Rescued-By-TSA-Agents-at-MIA-Authorities-164345996.html.

Sisson, Mary. “Richard Reid (British Militant).” *Encyclopedia Britannica Online*, March 5, 2014. http://www.britannica.com/EBchecked/topic/1975161/Richard-Reid.

Sullivan, Maureen, Mark G. Frank, Carolyn M. Hurley, and Jaspreet Tiwana. “Police Lie Detection Accuracy: The Effect of Lie Scenario.” *Law and Human Behavior* 33, no. 6 (2009): 530–38. doi: 10.1007/s10979-008-9166-4.

TSA. “Layers of Security.” Lasted revision July 23, 2014. http://www.tsa.gov/about-tsa/layers-security.

———. “Screening Passengers and Property.” https://www.tsa.gov/sites/default/files/assets/pdf/49_usc_chapters_401_to_501.pdf.

———. “Visible Intermodal Prevention and Response (VIPR).” August 23, 2013. http://www.tsa.gov/about-tsa/visible-intermodal-prevention-and-response-vipr.

United States Government Accountability Office. *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, DC: U.S. Government Accountability Office, 2010.

———. *Testimony Before the Subcommittee on Transportation Security, Committee on Homeland Security, House of Representatives Aviation Security: TSA Should Limit Future Funding for Behavior Detection, Statement of Stephen M. Lord, Director Homeland Security and Justice* (GAO-14-158T). Washington, DC: U.S. Government Accountability Office, 2013.

———. *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, DC: U.S. Government Accountability Office, 2010. http://www.gao.gov/new.items/d10763.pdf.

———. *Report to Congressional Requesters, Aviation Security: TSA Should Limit Future Funding for Behavior Detection Activities* (GAO-14-159). Washington, DC: U.S. Government Accountability Office, 2013.

———. *Transportation Security Administration: Progress and Challenges Faced in Strengthening Three Security Programs* (GAO-12-541T). Washington, DC: U.S. Government Accountability Office, 2012.

USAJOBS. "Transportation Security Officer (TSO)." Accessed October 7, 2014. https://www.usajobs.gov/GetJob/ViewDetails/383141100.

Vrij, Aldert. *Detecting Lies and Deceit: Pitfalls and Opportunities.* Chichester: John Wiley, 2008.

Walker, Estaban. "Meta-analysis: Its Strengths and Limitations." *Cleveland Clinic Journal of Medicine* 756 (June 2008): 431–39. doi: 10.3949/ccjm.75.6.431.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California