Behavioral and Social Indicators of Potential Violent Acts: Implications from a Review of the Science Base

Dr. Walter L. Perry
RAND Corporation
1200 South Hayes Street
Arlington, VA 22202
Tel: (703) 413-1100 Ext: 5228
Email: Walter_Perry@rand.org

Dr. Paul K. Davis
RAND Corporation
1776 Main Street
Santa Monica, CA 90407
Tel: (310) 393-0411 Ext: 6912
Email: Paul_Davis@rand.org

Dr. Ryan A. Brown
RAND Corporation
1776 Main Street
Santa Monica, CA 90407
Tel: (310) 393-0411 Ext: 6234
Email: Ryan_Brown@rand.org

Abstract

Governments have put substantial effort into thwarting terrorist attacks by observing suspicious behaviors of individuals. Although technologies and methodologies abound for contributing to such activities, the claims about effectiveness sometimes lack a clear basis in science and technology. This paper reviews the base in behavioral science for using new or nontraditional technology and methods to observe individual- or small-group behaviors that might help detect potential violent attacks. Five phases of threatening activity classes were identified: developing intent; preparation and ground-laying; immediate pre-execution; execution; and aftermath. Technology and methods allow for the detection of behavioral indicators that signal the presence of phase activities. Analysis can then assess whether the totality of information adds up to a basis for concern. Detection technologies and methods are examined for each of the activity indicators. Among the countless technological and analytic efforts, the most important fell into three cross-cutting classes defined by type of data: communication patterns; "pattern-of-life" data; and indicators relating to body movement and physiological state. Detection systems were characterized along seven dimensions; layering; behavioral stimulation; countermeasure resistance; the ability to obtain the information, perhaps surreptitiously; sensitivity and selectivity of subsystems with no stimulation; information fusion; and minimizing the consequences of false alarms. We draw several conclusions among which are: problems and errors can be avoided by detection and screening under conditions of high false alarm rates and low base rates; "operators" are often well ahead of the science base; information fusion is critical; and profound issues of privacy and civil liberties are raised by detection methodologies.

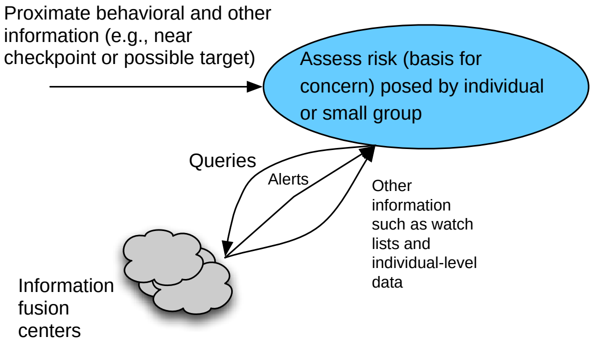
Introduction

In the United States, federal, state, and local government organizations have put substantial effort into detecting and thwarting terrorist and insurgent attacks by observing suspicious behaviors of individuals, whether at transportation checkpoints or in background activities such as participation in violent organizations. Technologies and methodologies abound for contributing to such defensive activities. The result, however, has often been overwhelming in its volume and the diversity of activities and claims. Further, the claims about effectiveness sometimes lack a clear basis in science and technology. RAND was asked to improve the situation by conducting an analytically useful review of the base in behavioral sciences relevant to threat detection, one that would help priorities for special attention and investment. [1]

Purpose and Approach

This study reviews the base in behavioral science for using new or nontraditional technology and methods to observe individual- or small-group behaviors that might—especially when used with other information—help detect potential violent attacks such as by suicide bombers or, as a very different example, insurgents laying improvised explosive devices (IEDs). Behavioral indicators may help identify individuals meriting additional observation in an operational context as depicted in Figure 1 in which security personnel are assessing (blue oval) whether an individual poses some risk in the limited sense of meriting more extensive and perhaps aggressive screening, follow-up monitoring, or intercept. The security personnel might be at a border crossing or watching a crowd gathered for a political speech. They obtain information directly, query data bases and information-fusion centers ("pull"), and are automatically provided alerts and other data ("push"). They report information that can be used subsequently. In some cases, behaviors of various individuals over time might suggest a potential ongoing attack even if the individuals are only pawns performing such narrow tasks as obtaining information.

Figure 1: Operational Context


Proximate behavioral and other information (e.g., near checkpoint or possible target)
Assess risk (basis for concern) posed by individual or small group
Queries
Alerts
Other information such as watch lists and individual-level data
Information fusion centers

The report is concerned with detecting imminent threats rather than gathering broad information for internal security or intelligence. However, some of the information used may have been accumulated over years. Where might that information be found, how might it be structured, and what indicators might be involved? Our report focuses on what may be possible technically, without analyzing tradeoffs with privacy and civil liberties. However, we note some of the troublesome issues raised by the technology and methods; further, we point readers to a study chaired by William J. Perry and Charles M. Vest (Perry and Vest (Chairmen), 2008), which included panelists from law, law enforcement, information technology, computer science, and other fields. Finally, we suggest some research on ways to mitigate the problems.

Figure 2 shows relationships among constructs. A base of technology and methods (left) allows detecting behavioral indicators (bottom right). Moving upward, these signal the presence of activities, which are lumped into activity classes called phases. Analysis can then assess whether the totality of information adds up to a basis for concern—i.e., more than usual justification for further screening, monitoring, precautionary defensive measures, or

[1] Another recent study relating to the prediction of violent behavior (Defense Science Board, 2012) raises and discusses many of the policy issues that we do not discuss here.

even preemptive action. Since detecting a “basis for concern” will probably have a high false alarm rate, a system using this approach must be very efficient and reasonable if it is to be acceptable.

Figure 2: Relationships Among Constructs

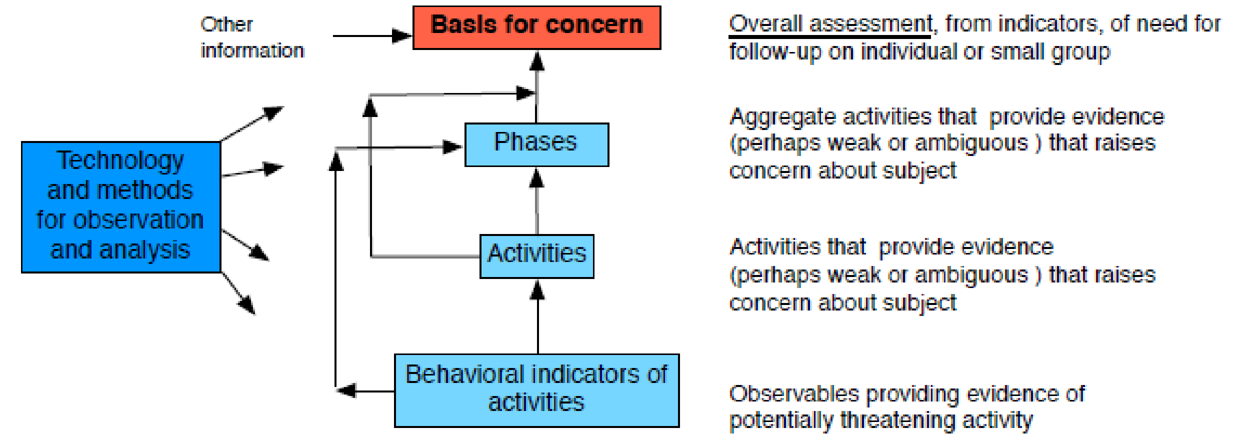

Other information
Basis for concern
Overall assessment, from indicators, of need for follow-up on individual or small group
Phases
Aggregate activities that provide evidence (perhaps weak or ambiguous) that raises concern about subject
Technology and methods for observation and analysis
Activities
Activities that provide evidence (perhaps weak or ambiguous) that raises concern about subject
Behavioral indicators of activities
Observables providing evidence of potentially threatening activity

Figure 3 is a “factor tree” conceptual model showing, along the top, phases within which lower-level activities occur.[2] The model merely identifies places to look for information. The activities may be of multiple individuals, some activities may not occur, or may occur in different order. A given activity can be associated with more than one phase. This diagram is neither a rigorous decomposition nor a timeline. As indicated at the bottom of the figure, there are many possible indicators of the activities and a number of technologies and methods to use in observing the indicators.[3]

[2] "Factor-tree" conceptual models were first used in earlier RAND studies and have proven quite useful in integrating and communicating heterogeneous social-science knowledge relating to terrorism, insurgency, and stabilization and reconstruction (Davis and Cragin, 2009; Davis, 2011; Davis et al., 2012). They can be seen as static simplifications of "causal-loop diagrams" or "influence diagrams" as used in system dynamics and policy analysis. The nodes (i.e., the factors or variables), need not, and typically are not, interpreted probabilistically, as are influence diagrams in Bayesian or influence-net research.

[3] Roughly analogous methods have been used in a variety of fields, such as with offender life cycles in criminology, where different phases are identified with respect to crime itself (one phase might be "search in a pre-criminal situation") and with respect to periods in a criminal's life, including a period of giving up crime. See, for example, the introduction chapter of Cornish and Clarke (1987). Process-model methods have been used to systematize counter proliferation research, as in identifying the numerous steps necessary to develop acquire, field, and employ a weapon of mass destruction. All such methods reflect one or another type of “system thinking,” some more rigorous than others.

Figure 3: Conceptual “Factor Tree” Model of Opportunities for Observation

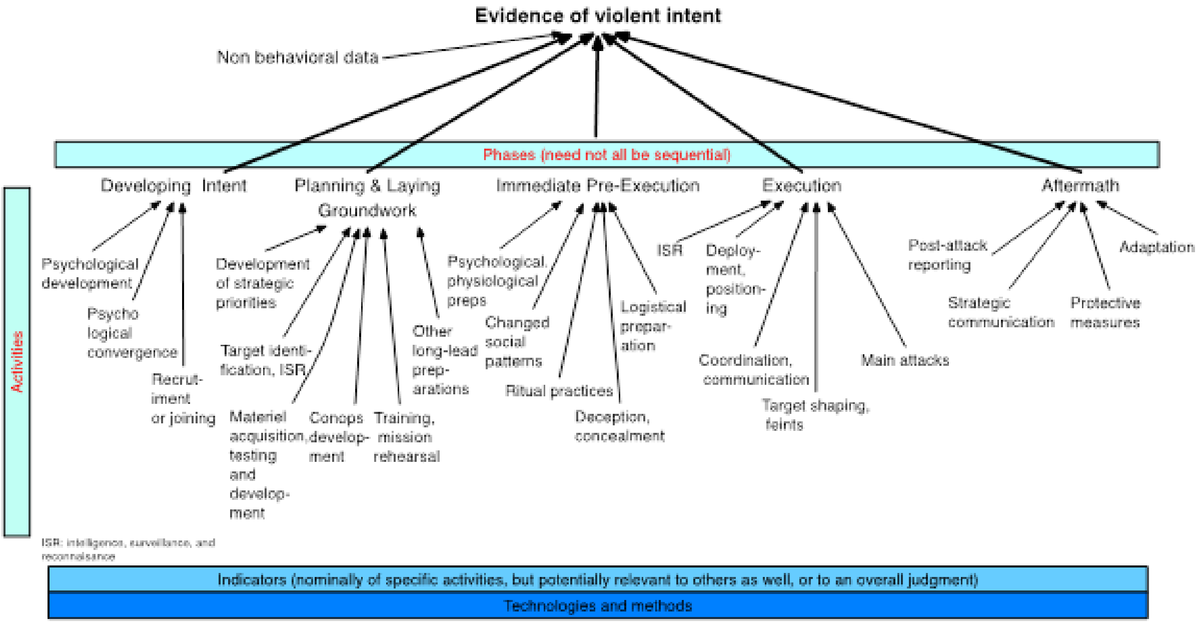
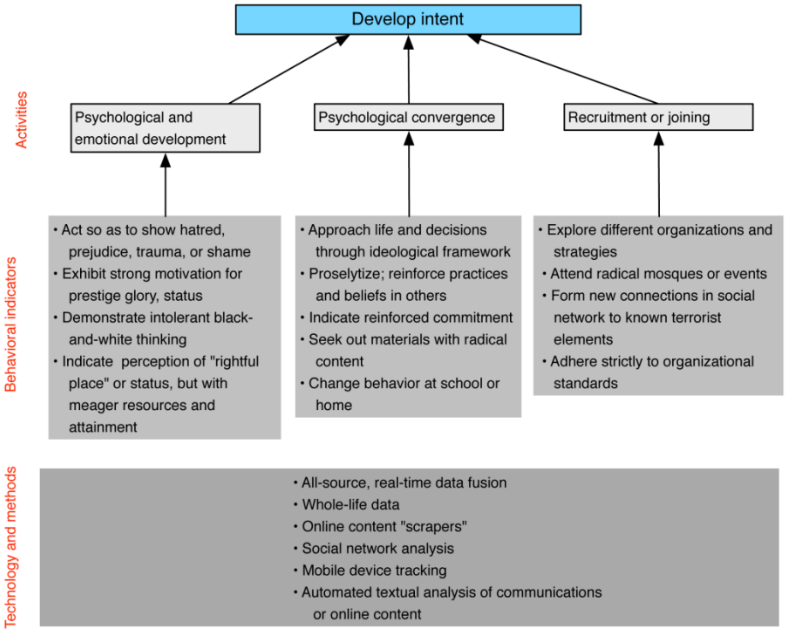

Evidence of violent intent
Non behavioral data
Phases (need not all be sequential)
Developing Intent
Planning & Laying Groundwork
Immediate Pre-Execution
Execution
Aftermath
Activities
Psychological development
Psycho logical convergence
Recrut-iment or joining
Development of strategic priorities
Target identi-fication, ISR
Materiel acquisition, testing and develop-ment
Concps develop-ment
Training, mission rehearsal
Other long-lead prep-arations
Psychological, physiological preps
Changed social patterns
Ritual practices
Deception, concealment
Logistical prepar-ation
ISR
Deploy-ment, position-ing
Coordination, communication
Target shaping, feints
Main attacks
Post-attack reporting
Strategic communication
Protective measures
Adaptation
ISR: intelligence, surveillance, and reconnaisance
Indicators (nominally of specific activities, but potentially relevant to others as well, or to an overall judgment)
Technologies and methods

Figure 4 illustrates the relationships among phases, activities, indicators, technologies and methods in our approach. It uses the example of the Developing Intent phase of Figure 3. For each of the three activities, it shows a number of potential indicators. In the lowest box, it shows some of the technologies and methods relevant to analysis. The Develop-Intent phase is unusual in that it includes early-in-life activities such as might be observed along the way by parents, neighbors, teachers, physicians, local law enforcement, and others—possibly long before an individual becomes involved in anything violent.

Figure 4: Illustration of Methodology


Develop intent
Activities
Psychological and emotional development
Psychological convergence
Recruitment or joining
Behavioral indicators
• Act so as to show hatred, prejudice, trauma, or shame
• Exhibit strong motivation for prestige glory, status
• Demonstrate intolerant black-and-white thinking
• Indicate perception of "rightful place" or status, but with meager resources and attainment
• Approach life and decisions through ideological framework
• Proselytize; reinforce practices and beliefs in others
• Indicate reinforced commitment
• Seek out materials with radical content
• Change behavior at school or home
• Explore different organizations and strategies
• Attend radical mosques or events
• Form new connections in social network to known terrorist elements
• Adhere strictly to organizational standards
Technology and methods
• All-source, real-time data fusion
• Whole-life data
• Online content "scrapers"
• Social network analysis
• Mobile device tracking
• Automated textual analysis of communications or online content

The Phases

The conceptual phase-level activities of Figure 3 are to some extent ambiguous and overlapping. We followed certain conventions in deciding in what phase a given activity or observable belongs. Overall, the primary issue is achieving approximate comprehensiveness, not cataloging each and every possible activity uniquely. This said, our rules-of-thumb conventions are as follows:

1. *Developing Intent.* This phase is associated more with individuals than with the organization; it is about motivation and commitment—whether to a cause, organization, or activity. An individual might be participating in organizational activities such as meetings or even general training, which have the effect of creating motivation and commitment. In contrast, previously motivated and committed individuals participating in the same activities might have their activities counted as part of the planning-and-laying-groundwork phase. The same observable activity might be listed in both phases.
2. *Preparation and Ground-Laying.* This phase is associated with both the individual and the organization[4] even though we may be observing individual behaviors. This is the phase in which the organization does its planning and prepares its people broadly for operations, perhaps with general physical training and the teaching of combat skills.
3. *Immediate Pre-Execution*. This phase, associated with the organization's perspective, is one in which plans are finalized and resources mobilized and maneuvered so as to make subsequent execution feasible, if decided upon. It might include increased reconnaissance (or, conversely, a period of reduced visibility because adequate information has already been obtained). It might mean deploying people to the relevant country, area, or city, but putting them in holding patterns.

[4] In the case of lone-wolf terrorists, the organization and individual are the same. The phases still apply, however.

4. *Execution*. This phase applies once a decision to commence the attack has been made (by the organization or, in a lone-wolf case, by the individual himself). This definition is consistent with the meaning of execution in military command and control. Execution may require initial activities such as maneuvering resources to their final attack locations (perhaps moving through or around checkpoints), final reconnaissance, arming of weapons, and coordination-related communications. Almost any execution operation is contingent, in that the attack can be called off along the way. Nonetheless, until and unless the attack is called off, activities in response to an "execute order" (or decision) are regarded as in the execution phase.
5. *Aftermath.* After an attack is accomplished or an in-process attack is terminated, activities are considered to be in the aftermath phase. This might include dispersing, vacating observation posts, pulling back agents, and communications related to escape or withdrawal.

Technology and Methods

Our literature survey revealed countless technological and analytic efforts, but we found that the most important fell into three cross-cutting classes defined by type of data: (1) communication patterns, (2) "pattern-of-life" data, and (3) indicators relating to body movement and physiological state.

Communication Patterns

Communications occurs in, e.g., face-to-face meetings, internet chat rooms, and cell phones. Large commercial and intelligence-sector investments have yielded techniques to monitor and analyze these communications, which we treat in three groups: online communications and analysis, text analysis and natural-language processing, and speech analysis.

Online Communication: Online statements and actions may reveal or suggest thoughts, emotions, or even intent. Thus, related tools and methods for analyzing online content and communications may be particularly helpful. Data collection I tself can be performed manually, but is more efficiently done using online-content "scrapers."[5] These can pull in content constantly from particular sites or individual authors, or can "flag" specific types of content.

Monitoring social media discussion for threatening communications is often the responsibility of human analysts, such as the New York Police Department's social media unit (Rock, 2011). The NYPD investigated threats posted on Twitter (e.g., "people are gonna die like Aurora") following the movie-theater shooting in Aurora, Colorado (Ruderman, 2012). Real-time social-media search tools can facilitate monitoring for discussions relating to potential violence. They may also track general discussion around such potential targets as landmarks, military bases, or upcoming events. Some large social media services, such as Twitter (Ruderman, 2012) and Skype (Timberg and Nakashima, 2012), have made content and user information available to law enforcement in the United States.

Keystroke loggers and other malware can reveal past and current searches for material; registrations or payments to training programs; location information; and information on past searches. Spikes or trends in activity may reveal when terrorists "go dark" before an attack; upticks may correspond to logistical preparations for an imminent attack or calls for vengeance after an event such as bin Laden's killing.

[5] Related terms include web harvesting and web data extraction. The technology is closely related to that for "indexing," which is central to the work of familiar internet search engines. Numerous scraper tools are readily available for download.

Some obvious shortcomings include false alarms (e.g., users may have benign reasons for their purchases or mere curiosity as they investigate troublesome web sites), low signal to noise ratio, and vulnerability to such counters as burying information amidst innocuous communication or using anonymous or false accounts (of which Facebook alone has 83 million). Finally, high-quality encryption is increasing as companies such as Apple increase security options available to developers and users for the iPhone (Garfinkel, 2012).

Text Analysis and Natural-Language Processing: Techniques to classify texts and analyze content are fairly well developed, although emotion and intent analysis are less so. Explicit content may include bragging, ideological statements, or admiration for terrorist leaders.[6]

Text analysis of *how* people write and talk can also shed light on thoughts and feelings. One prominent example is Linguistic Inquiry and Word Count 2007 (LIWC) (Pennebaker, Booth, and Francis, 2007; Pennebaker et al., 2007), simple word-counting software identifying word-usage patterns statistically associated with motivations, attitudes, emotions, and other psychological states. These patterns can be analyzed to extract topics of discussion (Pennebaker and Chung, 2008). Style of communication does not depend on a content specific topic and can suggest relationships and status within a social network. It has been used for, e.g., detecting corporate fraud, terrorist interrogations, and criminal testimony.

Natural-language processing can analyze massive amounts of text about which little is known in advance, classifying documents that can then be analyzed further by subject-matter experts. Clustering methods can identify concepts and such topics as weapons, tactics, or targets. Such mathematical techniques as latent semantic indexing can help understand concepts and have the advantage of being language independent. Machine translation can often turn foreign language texts into something analyzable without foreign-language expertise or language-specialized software. Speech-recognition technology can greatly increase the amount of text available for text analysis. It can also help identify individuals.

However, despite the considerable past research, further work will be required before— if ever—linguistic style analysis can be *reliably* used to detect deception. Noting that much prior work has used archival emails, a Deloitte report (Mosher, 2010) argues that LIWC-based deception research needs further testing and validation on “real-life data sets.” Similarly, in a national-academy review volume (Chauvin, 2011), the authors Chung and Pennebaker (2011), pioneers in such work, point out the need to adequately understand the perceiver/listener of potential deception, and the individual differences or situational factors that influence his/her judgment. In addition, detected patterns may be unrelated to any imminent threat and their interpretation often depends on cultural and individual idiosyncrasies.[7] A shortcoming of the research is that much linguistic-style analysis has been done only on archival data; much more testing and validation is needed with “real-life data sets.” Top researchers caution against expecting highly reliable detections or interpretations and suggest the need for very large data sets that reveal many cultural and individual differences.

[6] As one example, Major Nidal Hasan (the Fort Hood shooter) wrote a series of emails to Anwar al-Awlaki, subsequently released by the FBI. One referred to Hasan Akbar, an American Army soldier who killed two fellow soldiers and wounded 14 others in Kuwait in 2003. The email, reported by CCN said (Shaughnessy, 2012), with grammar errors retained:

> There are... many Muslims who join the armed forces for a myriad of different reasons. Some appear to have internal conflicts and have even killed or tried to kill other us soldiers in the name of Islam i.e. Hasan Akbar…
>
> Would you consider someone like Hasan Akbar or other soldiers that have committed such acts with the goal of helping Muslims/Islam (Lets just assume this for now) fighting Jihad and if they did die would you consider them shaheeds (martyrs)?

As often happens, the e-mail could be read at the time as not yet threatening.

[7] To illustrate with an example of cultural idiosyncrasies, a study of American and Japanese texts found that American authors used far more first person plural pronouns, in a distant, royal-we manner, as compared to Japanese authors (Fieldler, 2007).

Speech Analysis of Content: Several robust indicators exist for connecting vocal content and narratives with lying and deception. These include the subject: (1) distancing himself from untruthful statements by, e.g., using the third person or otherwise seeming less verbally involved; (2) issuing discrepant statements; (3) providing less detail; (4) exhibiting less logical structure and fewer subjectively plausible stories; (5) providing less context; and (6) making fewer spontaneous corrections or claiming lapses of memory (DePaulo et al., 2003).

Experiments with scenarios attempting to mimic potential terrorist attacks have shown positive results for some of these indicators—particularly subjective plausibility and the lack of consistency across statements or conversations (Vrij et al., 2011). Vrij and Granhag (2012) argue that vocal content in response to well-places probes or carefully crafted questions are the most reliable indicators of deception.

The approach's primary shortcoming in assessing deceptive or hostile intent is that interpreting lexical and vocal indicators of lying and deception depends on context, individual variability, and appreciation of non-threatening explanations. Optimally, analysis has the individual's data in a normal non-deceptive/non-hostile state. Where this is infeasible, there is the potential for increased failed detections and intolerably many false alarms. Table 1 summarizes results of our review for the assessment of communication patterns and content.

Table 1: Considerations and Caveats: Detection and Analysis of Communication Patterns

Domain	Status	Upside Potential	Measurement Requirements	Shortcomings and Vulnerabilities
Online communication and activities	Extensive collection and analysis occurs today for commercial and intelligence reasons. Technologies and methods for analyzing such online activities are still unproven in academic or operational settings.	Given trends, even more and varied interactions will be available for collection.	Tools already exist. However, challenges for dealing with massive volumes of noisy data are formidable.	• Methods have not been well validated in academic or operational settings • Low signal-to-noise ratio • Effects of encryption, using "code," using anonymizers, or moving offline.
Text analysis and natural-language processing	A considerable research base exists with numerous past applications. Even natural-language processing can be highly accurate in specific experimental settings.	Using operational data to train and to create baselines could improve detection of deception, hostility, or extremist patterns. Natural-language techniques, given training sets, could quickly analyze large amounts of data	Online text is naturally occurring and publicly accessible, requiring only passive collection. Active elicitation of text or oral statements is also possible in some security contexts such as checkpoints or interrogations.	• Context and cultural dependence • Inadequate testing in operational settings • Need for substantial data
Speech analysis: lexical and vocal cues	This has been validated in laboratory settings, including those specific to a counter-terrorism context	Advances in protocols for rapid assessment of speech patterns and content would have wide applicability for screening or other situations involving conversations with security personnel.	Such analysis currently requires interactions with security personnel asking questions, making it dependent on one-on-one efforts and judgments. For vocal tone, it requires relatively "clean" audio signals	• Physiological drivers such as anxiety and changes in vocal tone are individual-dependent • May be subject to counters, especially if criteria for judging, e.g., narrative credibility are known

Pattern-of-Life Data

It is possible to analyze patterns of communication, travel, purchasing, and other matters using existing records and databases, many of which are held by private industry. We discuss mobile-device tracking, using existing records, and machine learning for pattern detection. Before doing so, we should mention that there are profound social questions about what kind of data can and should be collected and analyzed.

Mobile-Device Tracking: Ubiquitous mobile devices provide a wealth of data on, e.g.: personal information, social relationships and networks, locations, patterns of movement and interactions,[8] preferences, political opinions,[9] the spread of information, and patterns of how opinions and preferences change. Smartphone usage data (Chittaranjan, Blom, and Gatica-Perez, 2012) are related to the Big Five personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism (Costa and Mccrae, 1990). These may provide some insight into motivations. Extroverted people are more likely to receive calls and spend more time talking.

As social networking through mobile devices becomes more commonplace, individuals are increasingly communicating with others that they never meet in person (Lampe, Ellison, and Steinfield, 2008). In some cases, relationships that occur entirely through mobile devices or online may be substituting for more "traditional" forms of social contact (Deresiewicz, 2011). The substitution is imperfect, however, and communication patterns and links derived from mobile-to-mobile communication is increasingly "muddy" and divorced from both intent to meet and intent to act in the offline or "real world." Clearly, this complicates drawing inferences about actual threat.

Existing Records: It is sometimes possible to develop individual profiles from information about, e.g., experiences, behaviors, and relationships over time, and to provide context for assessing other incoming data. The data could come from school records, criminal records, interrogation reports, and so forth. Additionally, surveillance cameras are now common in public and business settings allowing for the possibility of tracking an individual's pattern-of-life. Integrating such data requires analytic techniques, including those for all-source, real-time big-data fusion. Related analytic tools are increasingly available from such providers of cloud computing as Google and Amazon and social-media companies. Notably, however, commercial applications often do not require accuracy to improve the targeting of marketing efforts.

The shortcomings include, of course, the administrative, jurisdictional, legal, and database challenges of extracting and combining data across multiple sources and owners within and outside the Unites States.

Machine Learning and Big-Data Analysis: Given the sheer magnitude of data, it is increasingly important to analyze it without the benefit of prior hypotheses or known points of comparison. "Supervised" machine-learning techniques use known data sets to train the algorithms, which can then classify data, identify relationships, and discover concepts. "Unsupervised" learning analyzes data without the aid of such known comparisons. It seeks to find structure in the unlabeled data set. For example, researchers have used thousands of YouTube images for unsupervised detection of high-level features such as faces. Such techniques could be applied to learn and detect human bodies, and potentially, those suggesting imminent threat. Such machine learning techniques have been

[8] Analyzing millions of mobile call records, Kang et al. (2010) computed typical travel ranges at different times for individuals of different age and gender. Others have also estimated the predictability in people's whereabouts (Jensen et al., 2010) and future locations (Burbey and Martin, 2008).

[9] Madan and colleagues modeled individuals' exposure to diverse individuals and political information and how the diversity or lack of it affects opinion change (Madan, Farrahi, Gatica-Perez, and Pentland, 2011). Measuring diversity of information exposure and political opinion change may also suggest potential for identifying radicalization or Developing Intent activities.

applied to uncover fraud, to recognize deception in computer-mediated communication, and for predictive policing.[10] Artificial neural-network models are promising and can be applied in real-time. Video or image analysis and machine learning techniques could be employed to find, for example, such activities as shaving heads and praying activities in martyrdom videos.

One shortcoming is that machine learning techniques often require a large amount of data. At least in the public domain, sufficiently large databases of violent attacks and other events do not exist for topics such as terrorism. One innovative method for obtaining large, labeled data sets is to "crowdsource" the work of collecting and labeling individual pieces of information. The effectiveness of crowdsourcing has been demonstrated in other domains unrelated to terrorism, notably creating a dataset of emotional facial expressions (Mcduff, Kaliouby, and Picard, 2011).

Table 2, by analogy with Table 1, is our assessment of the various approaches focused on records-based whole-life information.

Table 2: Considerations and Caveats for Pattern of Life Data

Domain	Status	Upside Potential	Measurement Requirements	Shortcomings and Vulnerabilities
Mobile-device tracking	Algorithms to predict individual movement patterns, preferences, etc., have been developed and validated in laboratory and experimental settings, but can benefit from more naturalistic validation.	Mobile devices will continue to add connectivity features that enable tracking (e.g., location and motion sensors, Near Field Communication chips).	Mobile device tracking may require device-owner permissions or cooperation of communications network providers.	Such simple techniques as not traveling with a device or turning it off may defeat pattern-of-life algorithms based on mobile-device whereabouts. Mobile-to-mobile communication is often divorced from "real-life" behaviors and intent.
Pattern-of-life data	Validating techniques to analyze large amounts of pattern-of-life data may be difficult in controlled settings. Commercial data sets and analytic tools are increasingly available	Pattern-of-life data may allow integrating disparate data types to build fuller behavioral profiles on individuals of interest. Accessing and integrating data is an issue.	Measurement does not require active or voluntary consent. However, access to various databases held by commercial or private sources may be necessary.	Pattern-of-life data may be vulnerable to "cover" activities and behaviors. Databases and algorithms for detecting threatening patterns are in early development.
Machine learning and big-data analysis	Machine learning techniques have been extensively used and validated in experimental and some applied settings. Such techniques have been used in national security and law enforcement.	Machine learning and big-data analysis may "discover" unknown patterns or activities hidden in large amounts of data, but massive data is needed for training.	Measurement does not require active or voluntary consent. A large amount of data or a strong hypothesis regarding relevant activity is required.	Learning techniques are probabilistic and vulnerable to noisy data. Current systems do not understand how to associate behaviors of multiple threatening individuals.

[10] For example, data mining has been used to uncover fraud (Li, Yen, Lu, and Wang, 2012) and classification methods have been used to predict deception in computer-mediated communication (Zhou, Burgoon, Twitchell, and Qin, 2004).

Indicators from Physical Movement and Physiology

Behavioral science has identified a host of non-verbal behaviors associated with emotional and psychological state, and with deception and violent intent. These can be roughly categorized into (1) kinetics (including gross motor movements) and (2) observation of physiological state.

Kinetics and Gross Movement: Existing technology can collect data for kinematic patterns (movement). Surveillance and reconnaissance platforms (e.g., tower cameras or UAV systems) can monitor individuals as they maneuver before an attack. Video recording can view individuals before attacks and collect information on individuals who frequent potential attack sites, providing a baseline for identifying individuals engaged in pre-execution activities. For example, "gait signatures" may be compared against information in a database analogous to that of the controversial early-in-century DARPA program on Total Information Awareness (TIA) (Pugliese, 2011). Existing recordings of terrorism incidents (e.g., suicide bombings) may also provide baseline data for training new analysis tools. For example, Cohen, Morelli, and Scott (2008) proposed a method to model and flag potentially hostile intent gestures (including gait) from CCTV feeds for manual observation. This method, however, has yet to be tested experimentally.

Emotion— in targets as well as observers—plays a significant role in individuals' ability to detect or interpret gait or other body movements. People are most sensitive to detecting emotions associated with gait when the human walkers are expressing anger, as compared to walkers expressing other emotions or moving neutrally (Chouchourelou et al., 2006).

Incorporating emotion into machine-learning methods may increase their future utility. Affective computing may need to select from various psychology and neuroscience findings and theories of emotion (e.g., "appraisal models"). Often subsystems of monitoring and interpretation of stimuli can be computationally modeled. Improvements are possible when distinguishing between emotional states that differ in arousal, such as anger and sadness. Methods are being developed to analyze gait of people who may be carrying weighted objects, such as IEDs. These methods could be applied to such existing commercial technologies as cameras used for Kinect and Wii motion capture.

Observers may also be able to detect deceptive or clandestine movements. Research on deceptive motion has tended to focus on situations in which an actor attempts to deceive potential observers regarding the nature of his or her actions (e.g., pretending to lift boxes as though they were different weights (Runeson and Frykholm, 1983)) or whether people are truly performing an action (e.g., fake passing a ball vs. actually passing it (Kunde, Skirde, and Weigelt, 2011)). Research has shown that people are better at detecting deceptive movements if they are themselves experienced in those deceptive actions (Cañal-Bruland and Schmidt, 2009). In the 2009 Canal-Bruland and Schmidt study, veteran handball players and novices were asked to predict whether a simulated player shot or faked a shot. Skilled handball players significantly outdid novices in discriminating shots from fakes. People are also more likely to recognize intentionally deceptive actions by observing significant kinematics. However, observer expertise does not help determining deception when the body movement is incidental to the intended deception (Sebanz and Shiffrar, 2009):

> . . . studies have not investigated situations in which the body is consciously used as an instrument for deception. Rather, the focus has been on non- verbal signals that leak out without the individual's awareness…. Such a passive perspective on the body does not capture situations wherein movements are designed to be deceptive, such as when people fake injuries …

Analysis of kinetics and gross motor movements should apply to a wide variety of security contexts, although validation in naturalistic settings is needed and, as often occurs in looking for behavioral indicators, the indicators may arise for benign reasons, such as people being anxious at security screenings or checkpoints.

One challenge for gait analysis is that current detection systems and protocols are often built using simulated behaviors (e.g., with actors). More naturalistic (real-world) observations are needed.

Physiological State and Reactions: Observing physiological state and physiological changes holds promise for detecting deception and other behaviors. We touch upon polygraph testing, related use of peripheral nervous system response, use of electro encephalograms (EEGs), vocal stress, and facial-expression analysis.

The best-known approach to using physiological indicators is polygraph testing. It has been extremely contentious for decades, and continues to be.[11] The most definitive review was accomplished by the National Academy of Sciences in 2003 (National Research Council, 2003). Most work subsequent to the 2003 review has echoed or embellished the original findings, maintaining that physiological responses to conversational probes are highly context dependent and display dramatic variability within and across individuals, making their use questionable in courts of law (Vigluicci, 2009). Research has reiterated concerns about questioning techniques used during polygraph tests that may extract false confessions (Kassin et al., 2010; Porter and Ten Brinke, 2010). Only a small subset of nonverbal indicators is (weakly) correlated with lying (Vrij, 2010).

Despite these problems, enthusiasm for the methods continues in law-enforcement and intelligence communities, who argue that the methods are useful—as part of larger investigative processes,[12] which may deter lying, loosen tongues, and generate information (including confessions). In such investigations, the guilty party is also relatively likely to be among those tested, raising the "base rate." Polygraph methods, then, have proven value in forensic psychiatry (Grubin, 2010).

New technologies using electro encephalograms (EEGs) allow some physiological features to be observed without "wiring up" individuals, sometimes at a distance, and sometimes covertly or surreptitiously, as with using heat-sensitive cameras to detect capillary dilation and blood flow to the face and head. There is some evidence of unique value in indicating *deception or imminent action* by an individual *if* baseline information is available for that specific individual ahead of time and/or if credible intelligence about a possible attack is available. Most of the technologies are in a relatively early stage of development, but some seem to have potential. Measurement of physiological signals closer to the central nervous system (i.e., the brain) holds more promise for detecting guilt and behavioral intent. An example is the work of Meixner and Rosenfeld, which used electro-encephalograms (EEGs) to measure response to specific stimuli relative to the individual-specific baseline (Meixner and Rosenfeld, 2011).

Evidence of vocal tension and higher vocal frequency may also be predictors of stress and deception and a few observable aspects of speech are much more difficult for an individual to *control* than other indicators of deception (Villar, Arciuli, & Paterson, 2012). Vocal pitch and other non-lexical features of speech are measurable via a range of commercially available devices, each of which uses a different combination of frequency, pitch, and other parameters to assess vocal stress. Such techniques have been used by the DOD, Bureau of Prisons, Intelligence Community, and Law Enforcement agencies including the Los Angeles Sheriff's Department. They have reportedly not been embraced by the more cautious and skeptical intelligence communities (Pool, 2010, p. 11-12, referring to

[11] See the self-published Maschke and Scalabrini (2005) for a particularly harsh critique by authors who advocate against use of polygraphs.

[12] These larger processes might include nothing more than an extra round or so of questioning, or might involve trickery, psychological pressure, physical discomfort, and repeated rounds of interrogation. The national-academy study contains some case histories, which are illuminating, both positively and negatively (National Research Council, 2003).

discussion by Philip Ruben). An earlier review conducted for the Air Force Research Laboratory (Haddad et al., 2002) concluded that such methods—like polygraphs—can be useful in helping to obtain confessions during interrogations. Recent work, reviewing research over 30 years, concluded that the voice-stress technologies performed, in general, no better than chance (comments by Ruben in Pool, 2011, p. 11, citing work by Bhatt and Brandon that appears not to have been published in the public domain). Ruben went on to say (without citations) that questions exist about the underlying physiological hypotheses.

Much scientific work has concluded that humans share at least some universal facial expressions indicative of underlying emotional and motivational states (Ekman, 1970; Ekman and Rosenberg, 2005). While there has long been some academic disagreement,[13] particularly with respect to cultural differences (Scherer, 1970; Russell, 1995), proponents of universal emotions have defended with rigorous data analysis (Ekman, 1992a; Ekman, 1992b; Ekman, 1993). Cultural differences seem relegated to the secondary dampening or accentuation of emotional responses, the categorization and perception of emotional states (Jack et al., 2012), and enculturated "display rules" that cause minor differences in predominant facial expression tendencies determined by major facial muscle groups (Matsumoto, 1990). A strong expression of the science seems to be that:

> The seven fundamental emotions–anger, disgust, fear, happiness, sadness, surprise, and contempt–are displayed on the face with some fundamental features that are generally recognizable on all humans (barring neurological impairment).

For our purposes, the most promising domain of facial expression analysis is the detection of facial micro-expressions. Micro-expressions are involuntary expressions of fear, anger, or other emotions that display on the face for milliseconds, despite the best efforts of individuals to dampen or hide these expressions (Ekman, 2003). Whether the relevant behavior is early in the cycle of attack, or closer to the actual attack, facial micro-expressions potentially hold vital information about attackers and their intent. These micro-expressions can be detected via movements in the facial muscles that are coded as "action units."

At least currently, the two primary problems with using physiological indicators are (1) non-specificity (the indicators may stem from many causes, most of them benign); and (2) individual differences (the observables that indicate attack or deception differ markedly across individuals, which requires establishing sound individual-centered baselines). Countermeasures are a problem with polygraphs, but perhaps less so with EEG methods. Even with polygraphs, empirical results have varied. Some drugs, for example, have not reduced detection rates as expected, but physical training can be effective as a countermeasure. Controlling vocal stress indicators is difficult, but countermeasures can obscure distinctions between baseline and stressed behavior. Facial expressions suffer from the same problems of non-specificity, but have the advantage of being more closely linked to motivational state and intent than are other physiological signals. Individual differences are also important: a psychopathic attacker, for example, might be more inclined to show micro-expressions of "duper's delight" while passing through a checkpoint undetected, while a non-psychopathic attacker might instead show micro-expressions of fear (as would a perfectly harmless nervous traveler).

Table 3 is our assessment of how the approaches based on detecting intent from physiological indicators stand in terms of maturity, potential, measurability, and vulnerability to countermeasures.

[13] A popular-level summary of controversy regarding some of Ekman's work is Weinberger (2010), which draws on material from an unpublished JASON study and the claims of some of an inability to reproduce Ekman's work. For a journal-quality discussion, see Vrij and Granhag (2012) to the effect that the questioner-subject relationship is crucial and that assessment of unstimulated facial expressions (and other physiological observations) is ineffective. See Frank and Svetieva (2012) for a response. Controversy continues.

Table 3: Detecting Hostility or Deception from Movement Physiology and Movement

Domain	Status	Upside Potential	Measurement Requirements	Shortcomings and Vulnerabilities
Kinetics and gross Motor Movement	Indicators have been validated for human observation and automated analysis in laboratory and experimental settings, including some operational settings (e.g., for gait of individuals carrying weighted objects).	Gross motor movements may reveal action, intent, or deception. On-foot motions may be unavoidable in such proximal security settings as checkpoints. Gross motor movement may be passively observed, but also actively elicited.	Some security contexts may not allow for sufficient physical movement to be interpretable (e.g., interrogation).	• Masking with deceptive movements. • Sensitivity to context and individual differences. • Non-specificity: triggering by diverse emotions and motivations.
Physiological state and reactions	Indicators have been validated in laboratory and experimental settings, with some experimental paradigms simulating elements of counter-terrorism and some (facial) cutting across culture. In some cases (e.g., voice-stress and facial indicators), automated recognition shows potential, but currently has high error rates	Internal physiological reactions are relatively automatic and difficult to control (e.g., micro tremors in speech or micro facial expressions). Probing of various sorts (even seemingly random conversations) can trigger reactions. Certain elements of facial expression are very difficult to alter voluntarily, including micro-expressions	Currently, measurement requires direct application of sensors or the physical observation of, e.g., facial flushing, sweating, etc.) Some (e.g., facial) require lighting and proximity with currently painstaking coding feasible only for high-value interrogations). Success requires exceptional “natural” talent or training, but limited available data suggests training is effective. Measurements are most valuable when comparing to individual’s baseline, which is only feasible in voluntary monitoring or interrogation context.	• Differences across contexts and individuals. • Non-specificity. • Influence of drugs and training (e.g., to dampen or obscure differences between baseline and signals). • Masking, in some cases (e.g., sunglasses or plastic surgery) • Some differences exist (perhaps not critical) across culture. • Masking (e.g., sunglasses, plastic surgery, or Botox for facial), but this may also be an indicator)

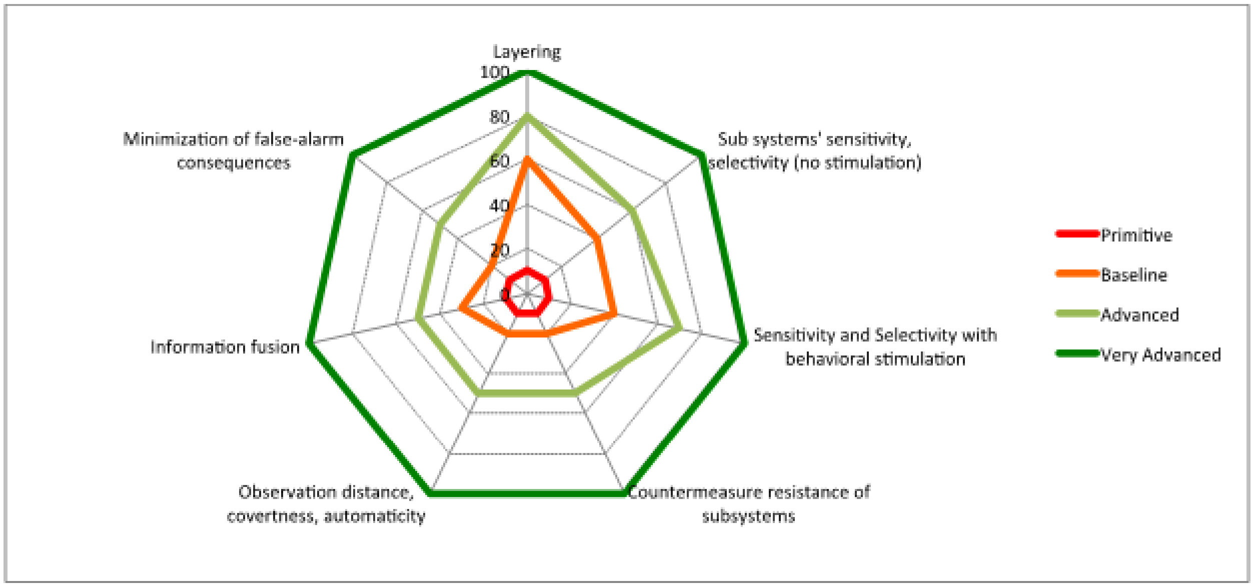
Cross-Cutting Themes

A number of cross-cutting themes arose in our review. Notionally, at least, these provide a kind of framework for thinking about detection systems. Although the relevant metrics have by no means been defined as yet, much less metrics that take into account cost-effectiveness, a goal for future analysis might be to place something like Figure 5 on a solid scientific and analytic basis. Although it is surely not yet "right" or well-defined, the figure may convey a sense of what is needed for sounder discussion. Further, despite all of its shortcomings, we have found the framework useful for discussing the issues that arose in our critical survey. That is, we found the *qualitative* framework useful.

Figure 5 characterizes a given detection system along seven dimensions, with a score of 100 corresponding to a system that has been optimized along that dimension, taking into consideration feasibility and cost effectiveness. The score given to a lesser system is a rough and subjective characterization of how much has been accomplished relative to what would be accomplished optimally. The dimensions relate to (1) layering; (2) behavioral stimulation; (3) countermeasure resistance of those subsystems; (4) the ability to obtain the information in desirable ways that may include automated observations from a distance, perhaps without subjects being aware of the observations; (5) sensitivity and selectivity of subsystems for information when it is obtained and countermeasures are absent; (6) information fusion; and (7) minimizing the consequences of false alarms when they occur.[14]

All of these dimensions relate to overall system effectiveness in detecting those that should be detected (minimizing "false negatives") and managing the false-alarm problem (minimizing "false positives" or their negative consequence). Figure 6 shows how we see the dimensions relating to overall objectives.[15]

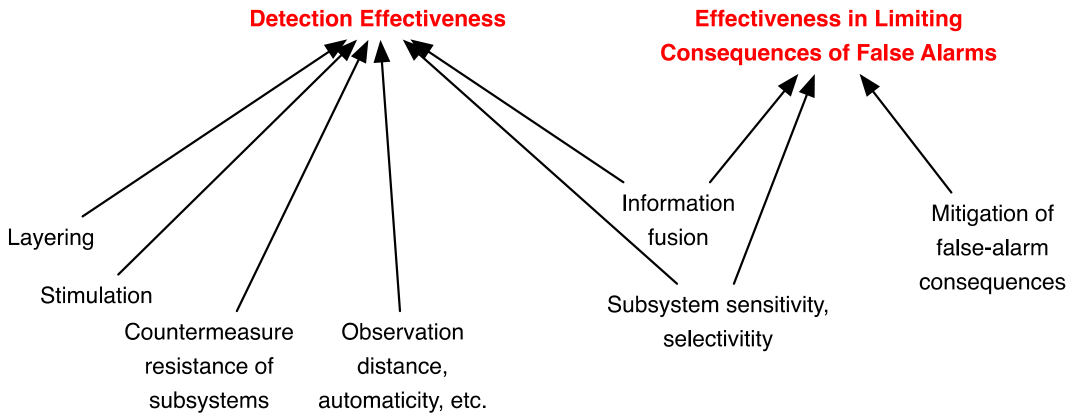
Figure 5: A Notional Framework for Characterizing the Overall System



[14] One absent theme is general data mining, such as collecting and mining behavioral data on all people in a population over time. Our focus is more on detecting attacks (e.g., at checkpoints or other defenses around targets) rather than, say, searching broadly for people with patterns of behavior that might relate somehow to terrorism, or in searching broadly for evidence that an individual is possibly subversive. Broad behavioral surveillance would raise especially profound issues of privacy, civil liberties, and the nature of pluralistic democracy (Perry and Vest (Chairmen), 2008).

[15] The area within a contour is not a sound measure of the option's overall effectiveness because not all of the dimensions are necessarily of equal importance and their value need not add linearly. The figure, however, is sufficient for our purposes in indicating that that progress involves progress along all of the dimensions shown.

Figure 6: Relationships To Overall System Effectiveness


Detection Effectiveness
Effectiveness in Limiting Consequences of False Alarms
Layering
Stimulation
Countermeasure resistance of subsystems
Observation distance, automaticity, etc.
Information fusion
Subsystem sensitivity, selectivitity
Mitigation of false-alarm consequences

Layering

Since no single fool-proof detector is plausible, good security-system designs exploit the potential leverage of layering, [16] i.e., of using a sequence of detection measures. Mathematically, the leverage can be dramatic. Suppose that an attacker must penetrate three independent layers, each of which has only a 1/3 chance of detecting him. With three layers, the defense has about a 70% likelihood of detecting him. The same benefits can be achieved by applying multiple detector methods at a given point in time, turning a single layer into the equivalent of multiple layers. A fatal flaw in some assessments is assuming that the various layers are independent, which is not the case when, for example, they share lax or incompetent management.

Subsystem Sensitivity and Selectivity

Screening can be based on many types of information, such as background checks, overtly observable characteristics (including the carrying of weapons), or behavioral cues. Ongoing research is a mix of laboratory- and field-based empirical research and modeling. As an example, the Department of Homeland Security's SPOT program was designed to provide behavior detection officers (BDOs) with a means of identifying persons who may pose a potential security risk at TSA-regulated airports in the United States. It focuses on behaviors and appearances that deviate from an established baseline and that may be indicative of stress, fear, or deception. The BDO officers may refer some of those whom they observe to additional screening. If behavioral indicators are used to classify people into in this way i.e., into those who are and are not regarded as representing more than normal risk, and should therefore be subject to further scrutiny, this can be referred to as negative screening—i.e., screening to identify people of concern because they do not "pass" all measures of "normalcy." Screening and the related issue of profiling[17] continue to be quite controversial, as discussed by the GAO (GAO, 2010), Congressional Research Service studies (Elias, 2009; Elias, 2011), and news media.

[16] The mathematics of layered defenses has been developed in the past for ballistic missile defense (Wilkening, 1999), defense in depth generally, and even cyber defense against worms (Albanese, Wiacek, Salter, and Six, 2004). A good treatment deals with countermeasures, game-theoretic considerations, and common-mode failures (Willis, Bonomo, Davis, and Hillestad, 2006); it discusses subtleties of inter-layer correlations (Jackson, La Tourette, et al., 2012, pp. 67-80 by La Tourette).

[17] The terms "screening" and "profiling" are not generally differentiated, but in some contexts "screening" is more descriptive and objective, while "profiling" *infers* characteristics, i.e., is more extrapolative. Sometimes, "profiling" refers to making decisions,

Screening/profiling based on or seemingly based on national origin, age, or apparent racial or ethnic grouping has generated some of the most heated debates (for a serious popular-level debate with informative discussion, see Harris and Schneier, 2012). Using a purely mathematical approach to analyze the advisability of racial or ethnic profiling, Press concludes that weak profiling (rather than what he calls democratic screening, which is when everyone is screened) is optimal (Press, 2009).[18] It is also is important to consider the secondary effects of such screening processes, as well as the possibility of malign actors thwarting simplistic screening procedures through the recruitment of operatives not displaying the screened characteristics (e.g., lack of an overtly Muslim or Middle Eastern appearance and use of female or child operatives).

A second class of screening method is illustrated by the "Trusted Traveler" concept, which screens for those who can be *excluded* from some secondary screening. Insights about the program were published early in the last decade (Shaver and Kennedy, 2004; Robinson et al., 2005). Robert Poole championed what he called a “risk-based” approach to screening (Poole and Passatino, 2003; Poole, 2009). The subject was reviewed analytically (Jackson, Chan, and Latourrette, 2011), assessing the value of using background checks to sort individuals into high and low risk categories for differential attention at checkpoints.

A constant issue in screening is how to trade-off detection rate against false-alarm rate. Doing so should depend on context. During a period of high alert, for example, security personnel can use less discriminate behavioral (and other) cues if they have additional temporary resources. During such a period, the public also tends to be more forgiving of inconvenience and somewhat higher false-alarm rates are tolerable.

Behavioral Stimulation

Probing to stimulate behavioral responses can sometimes improve detection effectiveness significantly. A recent review article expressed this as a general principle in using behavioral indicators, seeing it as key to progress (Vrij and Granhag, 2012). This view is consistent with long experience by Israel’s airport security personnel. However, the point can be exaggerated and the un-stimulated behaviors can sometimes be valuable in themselves or in combination (see also Frank and Syetieva, 2012, which responds to Vrij and Granhag). The basic concept has long been familiar to law enforcement and many tangible examples exist currently, partly as the result of U.S. security officials learning from extensive Israeli practice. More generally in our review, probing refers to the intentional stimulating of behavioral responses, perhaps by verbal questioning; anxiety-raising changes of procedure or process, subliminal stimuli, or tests with polygraph or EEG equipment. Probing may be polite, unobtrusive, or “in-your-face.” Some probing can definitely improve detection-system results, but related experimentation and formal assessment has not been pursued nearly as far as it might be. Verbal provocation and human assessment of verbal and behavioral responses can be quite effective in some circumstances without the use of sophisticated or expensive biological monitoring equipment. Israeli airport and other transit officials have used such techniques for many years, apparently with great success. Subjective assessment of the plausibility of reasons given for traveling, or being at a certain location, along with the consistency of stories over time together provide the best clues about hostile or deceptive intent. Further research should address contextually distinct tradeoffs between benefits for detection effectiveness and negative consequences for civil liberties, commerce, and the perceived legitimacy of the security system.

such as about whether to interrogate, based on racial or ethnic characteristics, which is illegal in many jurisdictions. Other times, "profiling" has no such negative meaning.

[18] Press defines "strong profiling" as screening in which the probability of being selected for secondary screening is at least proportional to a "prior" (the prior probability). A weaker version selects for secondary screening in proportion to the square root of the prior probability.

Allowing for and Dealing with Countermeasures

As noted in a recent study, “it should not immediately be assumed that the newest and most advanced technologies—the highest wall, the most sensitive surveillance—will best protect society from terrorist attack… it is only through fully exploring an adversary’s counter technology behaviors that vulnerabilities in a nation’s defenses can be discovered and the best choices made to protect the nation from the threat of terrorism” (Jackson et al., 2007, p.23). RAND has reviewed countermeasures taken by various groups to avoid detection—including Palestinian terrorist organizations, Jemaah Islamiyah, the Tamil Tigers, and the IRA. Documented countermeasures (including adaptations) include changes in patterns, style, and media for communication, disguise, false documentation, new weapons innovation, switching target sites, modifying attack duration, monitoring of monitoring devices and personnel, relying on the capabilities of more advanced affiliate organizations, destruction of forensic evidence, and punishment of informants to decrease HUMINT (Jackson et al., 2005). In each of these cases, the state security apparatus responded with its own adaptations too, including some that required technological development and refinement.

Much of the literature and even more of the advocacy-related discussion focuses on detecting behavioral responses in the absence of countermeasures, but countermeasures are in fact a big problem and vulnerability to countermeasures should be a prime consideration in evaluating investment programs. That said, countermeasures often are not employed, are attempted poorly, or themselves create indicators. Thus, a balance must be struck in analysis: worst-casing could eliminate valuable measures, but optimistic assumptions could waste resources and divert attention from more promising methods. Unfortunately, net judgments are often made informally and ad hoc. Improvements are possible.

Observation Circumstances: Remoteness, Covertness, Automaticity

Many of the technologies and methods with some effectiveness or considerable potential depend on relatively benign circumstance such as close-up observation by humans, sometimes with a need to minimize "noise" relevant to detection systems. Many tools exist for dealing with electronic information, but fewer tools exist for gait analysis than for data collection. DARPA has funded some biometric technologies, including HID (human identification at a distance) and VEW (video early warning) projects (Pugliese, 2008). Operational value will depend or be much enhanced by improved capabilities to make observations from a distance, automatically, and in some instances without the subjects being aware of the observation. Progress is being made by active technology efforts on all of these. Some of the efforts are benefiting from commercial and law-enforcement-system investments in, e.g., ubiquitous security-video recordings, supervised and unsupervised computer search of data, including "big data," and new analysis techniques such as those used in data mining.

The Potentially Critical Role of Information Fusion

A crucial step in assessing possible malign intent is reaching an overall assessment based on combining diverse indicators and overlaying inference (since the indicators are not able to reliably and selectively detect intent in most cases). The best that can be hoped for is a stronger sense of relative likelihood (even if small), so as to know when to look harder at an individual, or even take preemptive action. To illustrate, suppose that an individual attends one radical meeting a week for six months and also accesses radical websites. How likely is it that he is contemplating membership in a radical group? How do we combine knowledge from the two indicators? How do we “fuse” the indicator reports to modify our likelihood assessment? In this context, fusion is the process of combining information from various sources (similar and disparate in character) with the intention of obtaining a better composite of that being studied (see a review discussion in Perry, Signori, and Boon (2004)). Fusion may be accomplished with simple methods or much more sophisticated mathematical processes.

Heuristic and Simple-Model Methods include checklists and risk indexes, which are especially suitable for on-the-scene security personnel rather than, say, those in a fusion center. Checklists are common already and can be either negative (any indicator, if met, triggers additional screening) or positive (e.g., if all indicators are met, secondary screening can be minimized). Index methods (scoring methods) typically characterize a risk level by summing indicator scores, or by computing a risk as the product of a likelihood and a consequence, with a score exceeding a threshold triggering additional screening. [19] Scoring methods were used for decades in the Defense Department's force planning (Kugler, 2006). Score-based methods can be simplistic, moderately simple, or sophisticated, as with the Analytic Hierarchy Process (AHP) (Saaty, 1999). Significantly, good scoring methods often need to be nonlinear and should be empirically validated rather than ad hoc. We also consider more complex "simple methods" such as scorecards and conditional-indicator sets. [20]

More sophisticated integration methods are likely necessary (at least in future information-fusion centers, which may be quite different from today's) in using behavioral indicators because of serious signal-to-noise and false-alarm problems. [21] Accordingly, we reviewed in some detail the mathematical information-fusion methods that might be adapted and extended. Bayesian updating is now well understood and widely applied in other domains, but its usefulness in our context is limited by its demands for many subjective estimates of conditional probabilities for which there is and will continue to be an inadequate base (Feller, 1950; Mood and Graybill, 1963; Raiffa, 1968; Stone, Barlow, and Corwin, 1999). Some relatively new methods are based on Dempster-Shaefer belief-functions, which distinguish between having evidence *for* a proposition (such as the mal-intent of someone observed) and having contrary evidence (i.e., of innocence) (Shafer, 1976 and Shafer and Pearl, 1990b). Both can be high, whereas if the language used were that of simple probabilities, a high probability of mal-intent would imply a low probability of innocence. Depmster-Shaefer theory requires fewer subjective inputs. Ultimately, however, there are several major shortcomings in using that approach as well. A much newer approach, called Dezert-Smarandache (DSmT) theory has not yet been widely discussed and applied, but something along its lines has promise for our problem area because it deals specifically with combining evidence from sources and sensors that produce, imprecise, fuzzy, paradoxical and highly conflicting reports—precisely the type of reports encountered (Smarandache and Dezert, 2009b; Smarandache and Dezert, 2009a). For example, it allows characterizing the evidence that: both A and B are true; that one or the other of A or B is true (but not both); or the evidence that A is true and the evidence that A is not true. We also reviewed, briefly, the relevance of "possibility theory," (Dubois and Prade, 1988; Dubois and Prade, 1994) various multi-attribute theories, "mutual information" (which builds on the concept of information entropy), and Kalman filtering. The best method(s) for our problem area are not yet certain, but our review may help to generate fruitful research in this critical area.

[19] As an example, one such NIH index asks for age, gender, total cholesterol, HDL cholesterol, smoking (yes/no), systolic blood pressure, and medication for high blood pressure (yes/no). It then reports the likelihood of a heart attack over the next ten years (see http://hp2010.nhlbihin.net/atpiii/calculator.asp). The underlying formula is based on statistical analysis of medical data over many years.

[20] As noted in a classic paper by psychologist Robin Dawes, simple, and even linear decision aids have a track record of being remarkably effective in comparison with expert predictions (Dawes, 1979).

[21] Existing DHS Fusion centers have recently been discussed in scathing terms in a bipartisan Senate report by Senators Carl Levin and Tom Coburn. This report's references to fusion centers have in mind future centers that would have very different classes of information and analytic tools available to them. We did no research on the current centers. The Senate report is at http://www.hsgac.senate.gov/subcommittees/investigations/media/investigative-report-criticizes-counterterrorism-reporting-waste-at-state-and-local-intelligence-fusion-centers.

Mitigating Costs of False Alarms

As mentioned repeatedly, a major challenge in detection systems is the trade-off between false negatives (failure to detect) and false positives (false alarms), known as Type I and Type II errors. An understudied problem amenable to research is how the broadly construed cost of the latter can be reduced—not just by reducing the false-alarm rate, but also by mitigating the bad consequences of false alarms—consequences when people's time is wasted, their fears raised, their dignity insulted, or their privacy invaded (which, in turn, has bad societal effects). We identify three classes of initiative: (1) improve system effectiveness; (2) reduce effects of dignity and perceived violations of civil liberties (e.g., by transparency, explanation, fairness, apology, and compensation); [22] and (3) deter abuse by those within the security system. Progress on the latter two is highly desirable for broad societal reasons, and has many precedents in law enforcement. The negative consequences of false alarms alienate people, who are then less likely to cooperate, volunteer suspicions, and support the security system's mission.

A Core Issue in the Use of Behavioral Indicators

Many of the subjects reviewed in our study are extremely contentious. Some of the controversy is scientific, relating to whether various detection methods are scientifically sound (or, as some would have it, pseudo-science). The issue is not straightforward to discuss, however, because the critical task of detecting attacks by subjects such as terrorists involves looking for weak signals amidst a great deal of noise in circumstances in which the "base rate" is extremely low. The consequences of detection failure are very high, but so also there are profound negative consequences related to false alarms, some of which involve privacy, civil liberties, justice, and the pursuit of commerce and everyday life.

We could not hope to resolve the many controversies in this study, but Table 4 makes distinctions useful in discussion. It compares how various methods that use behavioral indicators can be used. All of them have deterrent or cost-imposition value (2nd column). Would-be attackers often fear the technology and methods and behave accordingly. All of the methods can, when properly used and in proper circumstance, be useful in providing some incremental evidence on which subjects merit closer scrutiny (3d and 4th columns), although there are big variations in whether they can be used automatically, remotely, and covertly. All of the methods, if well used, can *sometimes* (5th column) justify treating an individual with considerable concern, with subsequent assessment done "with prejudice" in the sense of being potentially, extended and including detention and aggressive questioning. That "sometimes" should be understood as "occasionally," however, and the methods typically have high false-alarm rates. Column 5 indicates that Yes, if a subject merits in-depth interrogation, all of the methods can—as part of a more complex process with skilled security officers—be useful in obtaining confessions or information. Regrettably, they can also help generate false confessions. The last column is crucial: none of the methods, except possibly for analysis of textual or vocal *content* (not really a behavioral method), are individually an adequate basis for arrest or conviction. Indeed, they may not be an adequate basis for putting prejudicial information in a widely shared data base (e.g., "On such-and-such an occasion, the subject manifested facial-expression behaviors correlated with posing a security risk, although other factors led to his being allowed to board the aircraft.”). This illustrates one of the many unresolved dilemmas. From a purely detection perspective, and assuming a process for information fusion, it would seem desirable to collect and share all kinds of fragmentary information of varied significance and credibility. However, doing so could cause serious injustices to those affected and, in many instances, would generate suspicions when none is scientifically warranted. It is instructive that, for almost a century, the FBI has maintained

[22] See a report from the Congressional Research Service for issues arising in airport screening and some efforts to allay concerns or mitigate issues (Elias, 2011). See also Helmus, Paul, and Glenn, (2007).

"raw files" on numerous subjects of observation, with some (but relatively few) instances of those files being misused. How much more trouble would have been created if that raw data had been widely shared? Arguably, such issues are matters of degree, but no common agreement exists on what is and is not reasonable. Interestingly, we all read routinely in the news media about how individuals in the news have allegedly passed or not passed lie detector tests over time, which surely must be prejudicial to the justice system. As one last example motivated by current discussions in the news (as of January 2013), consider a teenager with symptoms of schizophrenia being treated. What symptoms of violent tendencies should trigger a report to authorities that would enter a sharable data base, and with what balance of positive and negative consequences? Such issues are profound. We made no attempt to resolve them except that we see a major distinction between, on the one hand, using a behavioral indicator as an increment of information in a detection system seeking to identify, without further prejudice, which individuals merit more-than-usual scrutiny, and, on the other hand, inferring probable guilt or as the basis for arrest and conviction. It is not accidental that the U.S. justice system has major constraints on how methods such as polygraph techniques can be used.

Table 4 Some Comparisons of Behavioral Methods

Method	**Deterrence or Cost Imposition?**	**Flagging for further routine screening**		**Flagging with prejudice for extended checking and detention?**	**Tool in interrogation?**	**Basis for arrest or conviction?**
		Automatic?	**Human?**			
Polygraph	Yes	No	Yes	Maybe	Yes, but	No
VSA	Yes	Yes	Yes	No	Yes, but	No
Facial Expression	Yes	Technology not well developed	Yes	No	Yes, but	No
EEG	Yes	Technology not developed	Yes	Maybe	Yes, but	No
Text or Speech Content	Yes	Maybe	Yes	Maybe	Yes, but	Maybe
Gait Analysis	Yes	Yes	Yes	Maybe	No	No

Conclusions

We found a number of important "takeaways" from our survey:

- Despite the exaggerations found in commercial claims, studies, and the media, there is current value and unrealized potential for using behavioral indicators as *part* of a system to detect attacks. Unfortunately, analytic quantification of that potential is poorly developed.
- "Operators" are often well ahead of the science base, which is sometimes good and sometimes bad. It is very important that programs build in and sustain objective evaluation efforts, despite budgetary pressures and the tendency to see them as mere nice-to-have items. The evaluations should be subjected to objective peer review and adequate community scrutiny, even if security considerations would mean that should be accomplished within a domain of cleared personnel, limited distribution, etc. For example, FFRDCs,

National Academy, and other special national panels have conducted ranalogous evaluations for decades on a classified basis.

- Many serious problems and errors can be avoided by up-front review of procedures by experts familiar with the subtleties of detection and screening in conditions of high false alarm rates and low base rates. Although full validation of techniques may take years (at a time when the dangers of attack are current), avoiding many problems can be accomplished with existing knowledge. Some problems so avoided are quite significant to privacy, civil liberties, and the efficiency of travel and commerce.
- DHS and other security organizations are making efforts to experiment with and evaluate proposed methods—sometimes with laudable and ambitious scientific trials that have reported encouraging conclusions (which are difficult to judge without detailed access to data and methods).
- Operators, their agencies, and the scientific community have not done enough to understand how to mitigate the considerable bad consequences of detection systems, which invariably have false-alarm problems. Much could be done.
- Information fusion is critical, not just desirable, if behavioral indicators are to achieve their potential. Fusion should occur not just within a given method (as within polygraph methods), but with information across activities and phases. Methods for accomplishing this are very poorly developed. Also, it remains to be seen how much can realistically be accomplished.
- Information generation, information retrieval, integration, and sense-making will place enormous demands on both automated methods (e.g., including for "big data") and perfecting human-machine interactions: machines can process vast amounts of data, but interpretation will continue to depend critically on human expertise and judgment. “Optimizing” should be for man-machine cooperation, not automation.
- Very little research has been done to understand how much is enough, or what the curve of diminishing returns looks like, but, subjectively, it seems that *major* improvements in detection are plausible with networked real- or near-real-time integration of information. This would include not just integrating information of the CIA and FBI (much discussed since 9/11), but also in integrating (fusing) (a) proximate information at checkpoints with fusion-center information and (b) criminal, commercial, security-related, and even whole-life information. All of this is hypothesis. Developing a sharper understanding of payoff potential should be a priority task for objective research and analysis.
- Contemplating such steps raises profound issues of privacy and civil liberties, but the irony is that commercial organizations (and even political parties) are already far ahead in exploiting the relevant technologies and forever changing notions of privacy.
- Investment decisions about individual technologies and methods should be informed by a structured portfolio-analysis approach with criteria that include whether a given approach (1) applies to different types of threat and different activity classes for each (developing intent…execution and aftermath); (2) requires some kind of active stimulation; (3) applies to observation at a distance; (4) can discriminate between, say, anger or deception on the one hand, and many other emotions on the other; (5) can appropriately tailor interpretations to context or individual; or (6) would improve the data bases on which both methods and interpretations depend.

References

Albanese, David J. et al. (2004), *The Case for Using Layered Defenses to Stop Worms,* Fr. Meade, Maryland: National Security Agency.

Cañal-Bruland, R, and M Schmidt (2009), "Response Bias in Judging Deceptive Movements," *Acta psychologica*.

Chauvin, Cherie (for BBCSSS), (ed.) (2011), *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*, Washington, D.C.: National Academy Press.

Chittaranjan, Gokul, Jan Blom, and Daniel Gatica-Perez (2012), "Mining Large-scale Smartphone Data for Personality Studies," *Personal and Ubiquitous Computing*.

Chouchourelou, Arieta et al. (2006), "The Visual Analysis of Emotional Actions," *Social Neuroscience*, 1(1), 63-74.

Chung, Cindy K., and James W. Pennebaker (2011) "Using Computerized Text Analysis to Assess Threatening Communications and Actual Behavior," in *Threatening Communications and Behavior: Perspectives on the Pursuit of Public Figures*, edited by Claire Chauvin, Washington, DC: The National Academic Press, 3-32.

Cohen, Charles J., Frank Morelli, and Katherine A. Scott (2008), "A Surveillance System for the Recognition of Intent Within Individuals and Crowds," In *Proceedings of 2008 IEEE Conference on Technologies for Homeland Security*, http://www.openskies.net/papers/Intent Recognition.pdf.

Cornish, Derek B. and Ronald V. Clarke (eds.)(1986), *The Reasoning Criminal: Rational Choice Perspectives on Offending*, New York: Springer-Verlag.

Costa, Paul T., and Robert R. McCrae (1990), "Personality Disorders and the Five-factor Model of Personality," *Journal of Personality Disorders*, J4(4), 362-71.

Davis, Paul K. et al. (2012), *Understanding and Influencing Public Support for Insurgency and Terrorism*, Santa Monica, Calif.: RAND Corporation.

Davis, Paul K., (ed.) (2011), *Dilemmas of Intervention: Social Science for Stabilization and Reconstruction*, Santa Monica, Calif.: RAND Corporation.

Davis, Paul K., and Kim Cragin, (eds.) (2009), *Social Science for Counterterrorism: Putting the Pieces Together*, Santa Monica, Calif.: RAND Corporation.

Dawes, Robyn M. (1979), "The Robust Beauty of Improper Linear Models," *American Psychologist*, 34, 571-82.

DePaulo, Bella M. et al. (2003), "Cues to Deception," *Psychological Bulletin*, 129(1), 74-118.

Deresiewicz, William (2011) "Faux Friendship," in *Acting Out Culture*, edited by James S. Miller, Boston: St. Martin's Press, 470-80.

Dubois, Didier, and Henri Prade (1988), *Possibility Theory—An Approach to Computerized Processing of Uncertainty*, New York: Plenum Press.

Dubois, Didier, and Prade, Henri, (1994) “Possibility Theory and Data Fusion in Poorly Informed Environments,” *Control Engineering Practice*, Vol. 25, PP 811-823.

Ekman, Paul (1970), "Universal Facial Expressions of Emotion," *California Mental Health Research Digest. Vol.*, 8(4), 151-58.

——— (1992a), "Are There Basic Emotions?," Psychological Review, 99, No. 3, 550-53.

——— (1992b), "Facial Expressions of Emotion: New Findings, New Questions," Psychological Science, 3(1), 34-38.

——— (1993), "Facial Expression and Emotion," American Psychologist, 48, No. 4, 384-92.

——— (2003), "Darwin, Deception, and Facial Expression," Annals of the New York Academy of Sciences, 1000(1), 205-21.

Ekman, Paul, and Erika L. Rosenberg (2005), "What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (Facs), Second Edition," *Series in affective science. New York, NY, US: Oxford University Press*, 639.

Elias, Bart (2011), *Changes in Airport Passenger Screening Technologies and Procedures: Frequently Asked Questions*, Washington, D.C.: Congressional Research Service.

Feller, William (1950), *An Introduction to Probability Theory and Its Applications*, New York: Wiley.

Fieldler, K., (ed.) (2007), *The Psychological Functions of Function Words*, Social Communication.

Frank, Mark G., and Elena Svetieva (2012), *"Lies Worth Catching Involve Both Emotion and Cognition," Journal of Applied Research in Memory and Cognition, 1, 131-33.*

Frank, Mark G., and Elena Svetieva (2012), *"Lies Worth Catching Involve Both Emotion and Cognition," Journal of Applied Research in Memory and Cognition, 1, 131-33.*

GAO (2010), *Aviation Security: Efforts to Validate Tsa's Passenger Screening Behavior Detection Program Are Underway, But Opportunities Exist to Strengthen Validation and Address Operational Challenges*, Washington, D.C.: General Accountability Office.

Garfinkel, Simson L. (2012), "The iPhone Has Passed a Key Security Threshold," *Technology Review*.

Grubin, Don (2010), "The Polygraph and Forensic Psychiatry," *American Academy of Psychiatry and*, 38 (4), 446-51.

Harris, Sam, and Bruce Schneier (2012), "To Profile Or Not to Profile: A Debate," http://www.samharris.org/blog/item/to-profile-or-not-to-profile and http://www.schneier.com/essay-397.html, (last accessed, July 25, 2012).

Helmus, Todd C., Christopher Paul, and Russell W. Glenn (2007), *Enlisting Madison Avenue: The Marketing Approach to Earning Popular Support in Theaters of Operations*, Santa Monica, Calif.: RAND Corporation.

Jack, Rachael E. et al. (2012), "Facial Expressions of Emotion Are Not Culturally Universal," *Proceedings of the National Academy of Sciences*.

Jackson, Brian A. et al. (2005), *Aptitude for Destruction, Volume 1: Organizational Learning in Terrorist Groups and Its Implications for Combating Terrorism*, Santa Monica, Calif.: RAND Corporation.

Jackson, Brian A. et al. (2007), *Breaching the Fortress Wall: Understanding Terrorist Efforts to Overcome Defensive Technologies*, Santa Monica, CA: RAND Corporation.

Jackson, Brian A., Edward W. Chan, and Tom LaTourrette (2011), "Assessing the Security Benefits of a Trusted Traveler Program in the Presence of Attempted Attacker Exploitation and Compromise," *RAND Working Papers*, (WR-855-RC).

Jackson, Brian A., Tom LaTourette, Edward W. Chan, Russell Lundberg, Andrew R. Morral, and David R. Frelinger (2012), *Efficient Aviation Security: Strengthening the Analytic Foundation for Making Air Transportation Security Decisions*, Santa Monica: RAND.

Jensen, Bjørn Sand et al. (2010), "Estimating Human Predictability From Mobile Sensor Data," *IEEE International Workshop on Machine Learning for Signal Processing*.

Kang, Chaogui et al. (2010), "Analyzing and Geo-visualizing Individual Human Mobility Patterns Using Mobile Call Records," *Geoinformatics*.

Kassin, Saul M. et al. (2010), "Police-induced Confessions: Risk Factors and Recommendations," *Law and Human Behavior*, 34(1), 3-38.

Kugler, Richard L. (2006), *Policy Analysis in National Security Affairs: New Methods for a New Era*, National Defense University.

Kunde, Wilfried, Stefanie Skirde, and Matthias Weigelt (2011), "Trust My Face: Cognitive Factors of Head Fakes in Sports," *Journal of Experimental Psychology Applied*, 110-27.

Lampe, Cliff, Nicole B. Ellison, and Charles Steinfield (2008), "Changes in Use and Perception of Facebook," *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, 721-30.

Li, Shing-Han et al. (2012), "Identifying the Signs of Fraudulent Accounts Using Data Mining Techniques," *Computers in Human Behavior*.

Madan, Anmol et al. (2011), "Pervasive Sensing to Model Political Opinions in Face-to-face Networks," *Pervasive Computing*.

Maschke, George W., and Gino J. Scalabrini (2005), *The Lie Behind the Lie Detector*, AntiPolygraph.org.

Matsumoto, David (1990), "Cultural Similarities and Differences in Display Rules," *Motivation and Emotion*, 14(3), 195-214.

McDuff, Daniel, Rana el Kaliouby, and Rosalind Picard (2011), "Crowdsourced Data Collection of Facial Responses," *Proceedings of the 13th international Conference on Multimodal Interfaces (ICMI)*, 11-18.

Meixner, John B., and J. Peter Rosenfeld (2011), "A Mock Terrorism Application of the P300-based Concealed Information Test," *Pscyophysiology*, 48(2), 149-54.

Mood, Alexander, and Franklin Graybill (1963), *Introduction to the Theory of Statistics,* New York: McGraw-Hill.

Mosher, Daniel (2010), *Linquistic Deception Theory: Is the World of E-discovery Ready?,* Deloitte Financial Advisory Services LLP.

National Research Council (2003), *The Polygraph and Lie Detection*, Washington, D.C.: National Research Council.

Pennebaker, James W. et al. (2007), *The Development and Psychometric Properties of Liwc2007,* Austin, TX: LIWC.net.

Pennebaker, James W., and Cindy K. Chung (2008) "Computerized Text Analysis of Al-Qaeda Transcripts," in *A Content Analysis Reader,* edited by K. Krippendorff, and M.A. Bock, Thousand Oaks, Calif.: Sage.

Pennebaker, James W., Roger J. Booth, and ME Francis (2007), "Linguistic Inquiry and Word Count: Operator's Manual," (last accessed, September 18, 2012).

Perry, Walter L., David A. Signori, and John E. Boon (2004), *Exploring Information Superiority: A Methodology for Measuring the Quality of Information and Its Impact on Shared Awareness,* Santa Monica, CA: RAND Corporation.

Perry, William J., and Charles M. Vest (Chairmen) (2008), *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment*, Washington, D.C.: National Academies Press.

Pool, Robert W. (2011), *Field Evaluation in the Intelligence and Counterintelligence Context: A Workshop Summary*, Washington, D.C.: National Academy Press.

Porter, Stephen, and Leanne ten Brinke (2010), "The Truth About Lies: What Works in Detecting High-stakes Deception?," *Legal and Criminological Psychology*, 15(1), 57-75.

Press, William H. (2009), "Strong Profiling is Not Mathematically Optimal for Discovering Rare Malfeasors," *Proceedings of the National Academy of Sciences*, 106(6), 1716-19.

Pugliese, Joseph (2008), "Biotypologies of Terrorism," *Cultural Studies Review*, 14(2), 49-66.

Pugliese, Joseph (2011), "Biotypologies of Terrorism," *Cultural Studies Review*.

Raiffa, Howard (1968), *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*, Addison-Wesley.

Robinson, William H., Jennifer E. Lake, and Lisa M. Seghetti (2005), *Border and Transportation Security: Possible New Directions and Policy Options,* Congressional Research Service (CRS).

Rock, Margaret (2011), "Nypd to Scan Facebook, Twitter for Trouble," (last accessed, September 12, 2012).

Ruderman, Wendy (2012), "Court Prompts Twitter to Give Data to Police in Threat Case," The New York Times, A14, accessed at http://www.nytimes.com/2012/08/08/nyregion/after-court-order-twitter-sends-data-on-user-issuing-threats.html, September 15, 2012.

Runeson, Sverker, and Gunilla Frykholm (1983), "Kinematic Specification of Dynamics as an Informational Basis for Person-and-action Perception: Expectation, Gender Recognition, and Deceptive Intention," *Journal of Experimental Psychology*.

Russell, James A. (1995), "Facial Expressions of Emotion: What Lies Beyond Minimal Universality?" *Psychological Bulleting* 118(3), 379-391

Saaty, Thomas L. (1999), *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World, New Edition 2001 (Analytic Hierarchy Process Series, Vol. 2),* RWS publications.

Sebanz, Natalie, and Maggie Shiffrar (2009), "Detecting Deception in a Bluffing Body: The Role of Expertise," *Psychonomic Bulletin & Review*.

Shafer, Glenn (1976), *A Mathematical Theory of Evidence*, Princeton, New Jersey: Princeton University Press.

Shafer, Glenn, and Judea Pearl, (eds.) (1990b), *Readings in Uncertain Reasoning*, San Mateo CA: Morgan Kaufman.

Shaughnessy, Larry (2012), "Hasan's E-Mail Exchange with al-Awlaki; Islam, Money and Matchmaking," CNN, July 20. http://security.blogs.cnn.com/2012/07/20/hasans-e-mail-exchange-with-al-awlaki-islam-money-and-matchmaking.

Shaver, Russell D., and Michael Kennedy (2004), *The Benefits of Positive Passenger Profiling on Baggage Screening Requirements,* Santa Monica, Calif.: RAND Corporation.

Smarandache, Florentin, and Jean Dezert, (eds.) (2009a), *Advances and Applications of DSmT for Information Fusion*, Rehoboth: American Research Press.

——— (2009b) "An Introduction to DSmT," in *Advances and Applications of DSmT for Information Fusion*, edited by Florentin Smarandache, and Jean Dezert, Rehoboth: American Research Press.

Stone, Lawrence D., Carl A. Barlow, and Thomas L. Corwin (1999), *Bayesian Multiple Target Tracking*, Norwood, Mass.: Artech House.

Timberg, Craig, and Ellen Nakashima (2012), "Skype Makes Chats and User Data More Available to Police," A1, accessed at http://www.washingtonpost.com/business/economy/skype-makes-chats-and-user-data-more-available-to-police/2012/07/25/gJQAobI39W_story.html, September 16.

Vigluicci, Vincent V. (2009), "Calculating Credibility: State V. Sharma and the Future of Polygraph Admissibility in Ohio and Beyond," *Akron Law Review*, 42, 319-54.

Villar, Gina, Joanne Arciuli, and Helen Paterson (2012), "Vocal Pitch Production During Lying: Beliefs about Deception Matter," *Psychiatry, Psychology, and Law*, 1-10.

Vrij, Aldert et al. (2011), "Lying About Flying: The First Experiment to Detect False Intent," *Psychology, Crime & Law*, 17, No. 7, 611-20.

Vrij, Aldert, (2010), *Detecting Lies and Deceit: Pitfalls and Opportunities (2nd Ed.)*, New York, NY, US: John Wiley & Sons Ltd.

Vrij, Aldert, and Pär Anders Granhag (2012), "Eliciting Cues to Deception and Truth: What Matters Are the Questions Asked," *Journal of Applied Research in Memory and Cognition*, 1, 110-17.

Weinberger, Sharon (2010), "Airport Security: Intent To Deceive?," *Nature International Weekly Journal of Science*, 465, 412-15., http://www.nature.com/news/2010/100526/full/465412a.html, last accessed January 23, 2013.

Wilkening, Dean (1999), "A Simple Model for Calculating Ballistic Missile Defense Effectiveness," *Science & Global Security*, 8(2), 183-215.

Willis, Henry H. et al. (2006), *Capabilities Analysis Model for Missile Defense (Cammd): User's Guide*, Santa Monica, Calif.: RAND Corporation.

Zhou, Lina et al. (2004), "A Comparison of Classification Methods for Predicting Deception in Computer-mediated Communication," *Journal of Management Information Systems*, 20(4), 139-65.