M. B. CENETA/AP

M. ANZUONI/REUTERS

Intent to deceive?

Can the science of deception detection help to catch terrorists?
Sharon Weinberger takes a close look at the evidence for it.

In August 2009, Nicholas George, a 22-year-old student at Pomona College in Claremont, California, was going through a checkpoint at Philadelphia International Airport when he was pulled aside for questioning. As the Transportation Security Administration (TSA) employees searched his hand luggage, they chatted with him about innocuous subjects, such as whether he'd watched a recent game.

Inside George's bag, however, the screeners found flash cards with Arabic words — he was studying Arabic at Pomona — and a book they considered to be critical of US foreign policy. That led to more questioning, this time by a TSA supervisor, about George's views on the terrorist attacks on 11 September 2001. Eventually, and seemingly without cause, he was handcuffed by Philadelphia police, detained for four hours, and questioned by Federal Bureau of Investigation agents before being released without charge.

George had been singled out by behaviour-detection officers: TSA screeners trained to pick out suspicious or anomalous behaviour in passengers. There are about 3,000 of these officers working at some 161 airports across the United States, all part of a four-year-old programme called Screening Passengers by Observation Technique (SPOT), which is designed to identify people who could pose a threat to airline passengers.

It remains unclear what the officers found anomalous about George's behaviour, and why he was detained. The TSA's parent agency, the Department of Homeland Security (DHS), has declined to comment on his case because it is the subject of a federal lawsuit that was filed on George's behalf in February by the American Civil Liberties Union. But the incident has brought renewed attention to a burgeoning controversy: is it possible to know whether people are being deceptive, or planning hostile acts, just by observing them?

Some people seem to think so. At London's Heathrow Airport, for example, the UK government is deploying behaviour-detection officers in a trial modelled in part on SPOT. And in the United States, the DHS is pursuing a programme that would use sensors to look at nonverbal behaviours, and thereby spot terrorists as they walk through a corridor. The US Department of Defense and intelligence agencies have expressed interest in similar ideas.

Yet a growing number of researchers are dubious — not just about the projects themselves, but about the science on which they are based. "Simply put, people (including professional lie-catchers with extensive experience of assessing veracity) would achieve similar hit rates if they flipped a coin," noted a 2007 report[1] from a committee of credibility-assessment experts who reviewed research on portal screening.

"No scientific evidence exists to support

the detection or inference of future behaviour, including intent," declares a 2008 report prepared by the JASON defence advisory group. And the TSA had no business deploying SPOT across the nation's airports "without first validating the scientific basis for identifying suspicious passengers in an airport environment", stated a two-year review of the programme released on 20 May by the Government Accountability Office (GAO), the investigative arm of the US Congress.

In response to such concerns, the TSA has commissioned an independent study that it hopes will produce evidence to show that SPOT works, and the DHS is promising rigorous peer review of its technology programme. For critics, however, this is too little, too late.

The writing's on the face

Most credibility-assessment researchers agree that humans are demonstrably poor at face-to-face lie detection. SPOT traces its intellectual roots to the small group of researchers who disagree — perhaps the most notable being Paul Ekman, now an emeritus professor of psychology at the University of California Medical School in San Francisco. In the 1970s, Ekman co-developed the 'facial action coding system' for analysing human facial expressions, and has since turned it into a methodology for teaching people how to link those expressions to a variety of hidden emotions, including an intent to deceive. He puts particular emphasis on 'microfacial' expressions such as a tensing of the lips or the raising of the brow — movements that might last just a fraction of a second, but which might represent attempts to hide a subject's true feelings. Ekman claims that a properly trained observer using these facial cues alone can detect deception with 70% accuracy — and can raise that figure to almost 100% accuracy by also taking into account gestures and body movements. Ekman says he has taught about one thousand TSA screeners and continues to consult on the programme.

Ekman's work has brought him cultural acclaim, ranging from a profile in bestselling book *Blink* — by Malcolm Gladwell, a staff writer for *The New Yorker* magazine — to a fictionalized TV show based on his work, called *Lie to Me*. But scientists have generally given him a chillier reception. His critics argue that most of his peer-reviewed studies on microexpressions were published decades ago, and much of his more recent writing on the subject has not been peer reviewed. Ekman maintains that this publishing strategy is deliberate — that he no longer publishes all
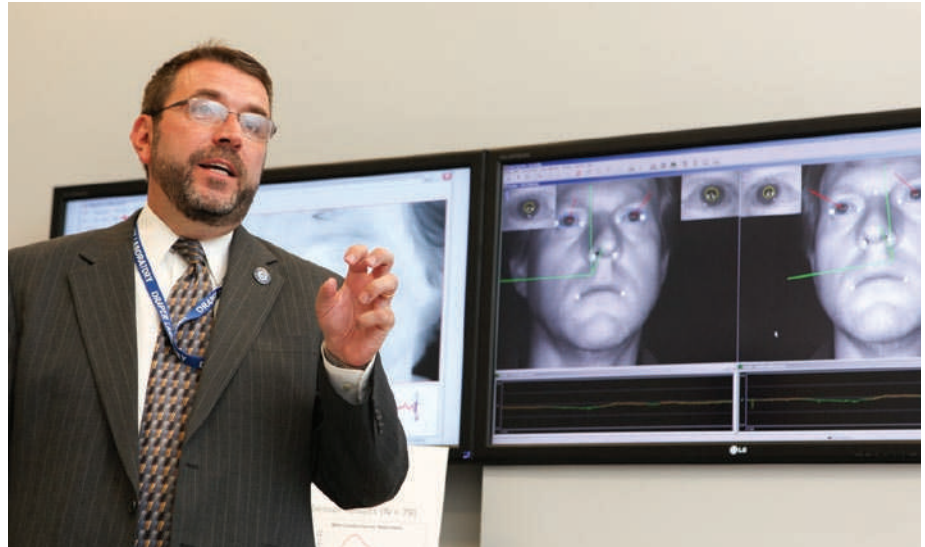
Robert Burns explaining the Future Attribute Screening Technology, which measures nonverbal cues.

of the details of his work in the peer-reviewed literature because, he says, those papers are closely followed by scientists in countries such as Syria, Iran and China, which the United States views as a potential threat.

The data that Ekman has made available have not persuaded Charles Honts, a psychologist at Boise State University in Idaho who is an expert in the polygraph or 'lie detector'. Although he was trained on Ekman's coding system in the 1980s, Honts says, he has been unable to replicate Ekman's results on facial coding. David Raskin, a professor emeritus of psychology at the University of Utah in Salt Lake City, says he has had similar problems replicating Ekman's findings. "I have yet to see a comprehensive evaluation" of Ekman's work, he says.

Ekman counters that a big part of the replication problem is that polygraph experts, such as Honts and Raskin, don't follow the right protocol. "One of the things I teach is never ask a question that can be answered yes or no," Ekman says. "In a polygraph, that's the way you must ask questions." Raskin and Honts disagree with Ekman's criticism, saying that Ekman himself provided the materials and training in the facial-coding technique.

Yet another objection to Ekman's theory of deception detection is his idea of people who are naturally gifted at reading facial expressions. These "wizards", Ekman argues[2,3], are proof that humans have the capability to spot deception, and that by studying those abilities, others can be taught to look for the same cues. But in a critique[4] of Ekman's work, Charles

"Linking displays of emotion to deception is a leap of gargantuan dimensions."

Bond, a psychologist retired from Texas Christian University in Forth Worth, argues that Ekman's wizard theory has a number of flaws — perhaps the most crucial being that the most successful individuals were drawn out of a sample pool in the thousands. Rather than proving these people are human lie detectors, Bond maintains, the wizardry was merely due to random chance. "If enough people play the lottery, someone wins," says Bond.

Ekman says that Bond's criticism is a "ridiculous quibble" and that the statistics speak for themselves. The wizards' scores were based on three different tests, he says, making it impossible to assign their high success rate to chance. Bond replies that he took the three tests into account, and that doing so doesn't change his conclusion.

Leap of logic

But there is yet another problem, says Honts. Ekman's findings are "incongruent with all the rest of the data on detecting deception from observation". The human face very obviously displays emotion, says Maria Hartwig, a psychology professor at the City University of New York's John Jay College of Criminal Justice. But linking those displays to deception is "a leap of gargantuan dimensions not supported by scientific evidence", she says.

This point is disputed by one of Ekman's collaborators, Mark Frank, a psychologist at the University at Buffalo in New York. Although Frank acknowledges that many peer-reviewed studies seem to show that people are not better than chance when it comes to picking up signs of deception, he argues that much of the research is skewed because it disproportionately involves young college students as test subjects,
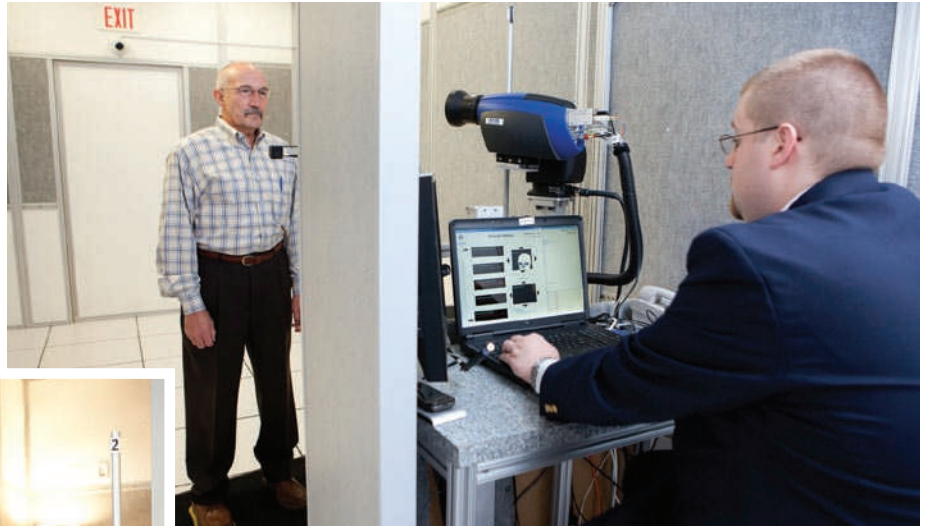
as opposed to police officers and others who might be older, more motivated and more experienced in detecting lies. Moreover, he says, when law-enforcement officials are tested, the stakes are often too low, and thus don't mimic a real-world setting. "I think a lot of the published material is still important, good work about human nature," says Frank. "But if you want to look at the total literature, and say, let's go apply it to counter-terrorism, it's a huge mistake."

A confounding problem is that the methodology used in SPOT, which is only partially based on Ekman's work, has never been subjected to controlled scientific tests. Nor is there much agreement as to what a fair test should entail. Controlled tests of deception detection typically involve people posing as would-be terrorists and attempting to make it through airport security. Yet Ekman calls this approach "totally bogus", because those playing the parts of 'terrorists' don't face the same stakes as a real terrorist — and so are unlikely to show the same emotions. "I'm on the record opposed to that sort of testing," he says.

But without such data, how is anyone supposed to evaluate SPOT — or its training programmes? Those programmes are "not in the public scientific domain", says Bella DePaulo, a social psychologist at the University of California, Santa Barbara. "As a scientist, I want to see peer-reviewed journal articles, so I can look at procedures and data and know what the training procedures involve, and what the results do show."

Carl Maccario, a TSA analyst who helped to create SPOT, defends the science of the programme, saying that the agency has drawn on a number of scientists who study behavioural cues. One he mentions is David Givens, director of the nonprofit Center for Nonverbal Studies in Spokane, Washington. Givens published a number of scholarly articles on nonverbal communications in the 1970s and 1980s, although by his own account he is no longer involved in academic research. His more recent publications include books such as *Your Body at Work: A Guide to Sight-Reading the Body Language of Business, Bosses, and Boardrooms* (2010). But Givens says that he has no idea which nonverbal indicators have been selected by the TSA for use in SPOT, nor has he ever been asked by the TSA to review their choices.

In the absence of testing, Maccario points to anecdotal incidents, such as the 2008 case of Kevin Brown, a Jamaican national who was picked out by behaviour-detection officers at Orlando International Airport in Florida and arrested with what they took to be the makings of a pipe bomb. Witnesses said that Brown was rocking back and forth and acting strangely, so it is hard to say whether specialized training was needed to spot his unusual behaviour. In any case, Brown successfully claimed that the 'pipe bomb' materials were actually fuel bottles, pleaded guilty to bringing a flammable substance onto an aircraft, and was released on three years' probation.

Cameras (above) and sensors (inset) can measure subtle physiological changes to eye movement, pupil dilation, heart rate and respiration, among other things.

Arrest record

The TSA does track statistics. From the SPOT programme's first phase, from January 2006 through to November 2009, according to the agency, behaviour-detection officers referred more than 232,000 people for secondary screening, which involves closer inspection of bags and testing for explosives. The agency notes that the vast majority of those subjected to that extra inspection continued on their travels with no further delays. But 1,710 were arrested, which the TSA cites as evidence for the programme's effectiveness. Critics, however, note that these statistics mean that fewer than 1% of the referrals actually lead to an arrest, and those arrests are overwhelmingly for criminal activities, such as outstanding warrants, completely unrelated to terrorism.

According to the GAO, TSA officials are unsure whether "the SPOT program has ever resulted in the arrest of anyone who is a terrorist, or who was planning to engage in terrorist-related activity". The TSA has hired an independent contractor to assess SPOT. Ekman says he has been apprised of the initial findings, and that they look promising. But the results aren't expected until next year. "It'll be monumental either way," says Maccario.

SPOT was in its first full year of operation when the DHS science and technology directorate began to look at ways to move people through the screening points faster. One was Future Attribute Screening Technology (FAST), which is now being funded at around US$10 million a year. The idea is to have passengers walk through a portal as sensors remotely monitor their vital signs for 'malintent': a neologism meaning the intent or desire to cause harm.

FAST operates on much the same physical principle as the century-old polygraph, which seeks to reveal lies by measuring psychophysiological responses such as respiration, cardiac rate and electrical resistance of the skin while a subject is being asked a series of questions. The FAST portal would also look at visual signals such as blink rate and body movement — and would give up the polygraph's contact sensors in favour of stand-off sensors such as thermal cameras, which can measure subtle changes in facial temperature, and BioLIDAR, a laser radar that can measure heart rate and respiration.

Most of the FAST work, particularly the

David Raskin.

F. MICELOTTA/GETTY IMAGES FOR FOX
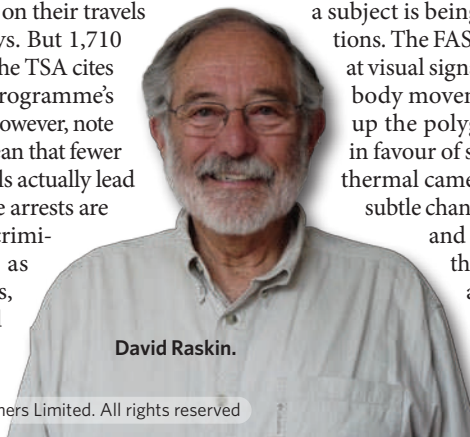
sensors, is contracted out to the Charles Stark Draper Laboratory, an independent, not-for-profit, research centre in Cambridge, Massachusetts, which has the goal of producing a prototype portal next year. The project is then scheduled to enter a second phase that will remove the questioning process altogether and instead try to induce a response in the subjects by using various stimuli such as sounds or pictures, possibly of a known terrorist. "In the laboratory now, we have a success detection rate [percentage] of malintent or not malintent, in the mid-70s," says Robert Burns, the DHS programme manager for FAST. "That's significantly better than chance or what the trained people can do."

Those results have not yet been published, but Burns says that the FAST programme sets great store on peer review and publication, and that three papers are currently in the process of review. But FAST's critics maintain that the malintent theory and FAST both suffer from some of the same scientific flaws as SPOT. Flying is stressful: people worry about missing flights, they fight with their spouses and they worry about terrorism. All of these stresses heighten the emotions that would be monitored by the FAST sensors, but may have nothing to do with deception, let alone malintent. "To say that the observation is due to intent to do something wrong, illegal or cause harm, is leaping at the Moon," says Raskin.

The malintent theory underlying FAST is the creation of Daniel Martin, who is the director of research for FAST, and his wife, Jennifer Martin. Both are psychologists, and Daniel Martin, who is on the faculty of Yale University in New Haven, Connecticut, has in the past focused primarily on the area of substance abuse. Daniel Martin says that at the time he and his wife developed the malintent theory, "there was minimal published work available that specifically tested whether physiological, behavioural, and paralinguistic cues could detect malintent in a realistic applied research study". He says that they have had to develop their own laboratory protocols to carry out those tests. Martin and his colleagues have just published what they say is the first peer-reviewed study to look specifically at the links between psychophysiological indicators and intent. The study[5] looks at 40 native Arabic-speaking men and finds a connection between intent to deceive and a heart-rate variation
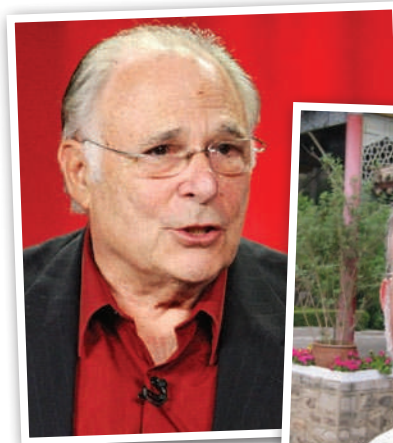
Paul Ekman (above) and Charles Honts disagree on deception detection.
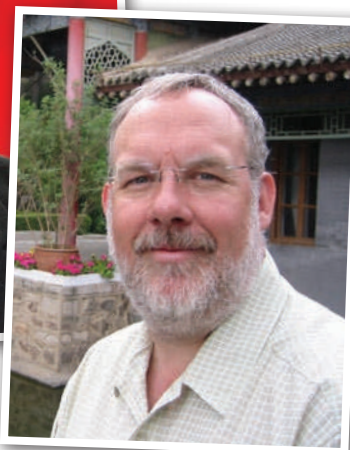
known as respiratory sinus arrhythmia.

"I have not come out and said, 'We have found the answer'," Martin adds. "We are pursuing the answer, we're not sure yet. We have years yet to go."

The lack of answers has not stopped aviation-security programmes from moving forwards with deception detection. Maccario points to the UK pilot scheme, now in its first year at Heathrow Airport. He says that the programme, like SPOT, uses specially trained behaviour-detection officers, and "their initial results are very successful". Earlier this year, the US Intelligence Advanced Research Projects Activity announced its own plans to study "defining, understanding, and ultimately detecting valid, reliable signatures of trust in humans". And about two years ago, the Pentagon asked JASON to look at the field.

"As we dug in, we found it was very hard to subject the research to the kinds of standard we're used to in the physical sciences," says JASON head Roy Schwitters, a physics professor at the University of Texas at Austin. In fact, the executive summary of the JASON report, *The Quest for Truth: Deception and Intent Detection*, which was provided to *Nature* by the Pentagon, criticizes many of the allegedly successful results from deception-detection techniques as being post-hoc identifications. One problem, the study found, was that the reported success rates often included drug smugglers, warrant violators and other criminals, not covert combatants or suicide bombers who might not have the same motivations or emotional responses.

Sallie Keller, dean of engineering at Rice University in Houston, Texas, and the head of the JASON study, said that it seemed that those involved in the field were trying to get their work peer reviewed. But doing research — even if it is properly peer reviewed — doesn't

"We are pursuing the answer, we're not sure yet. We have years yet to go."

mean the technology is ready to be used in an airport. "The scientific community thinks that it is extremely important to go through the process of scientific verification, before rolling something out as a practice that people trust," she says.

Researchers involved in the field suggest a number of research avenues that could be more fruitful for counter-terrorism. Aldert Vrij, a social psychologist at the University of Portsmouth, UK, says that structured interviews may offer the best credibility-assessment research. Nonverbal cues might play a part in this process, he says, but you need to actively interview a person. For example, his work shows that subjects were able to give more reasons for supporting an opinion that they believed than if they were acting as a devil's advocate and feigning support[6]. He suggests that such an approach could have helped to determine the beliefs of the Jordanian suicide bomber who killed seven CIA employees in Afghanistan after being taken into their confidence.

Although Israeli aviation security uses interview-intensive screening, it's not clear how practical such an interview method would be at busy airport checkpoints, which have to screen hundreds or thousands of passengers every hour. The guards would still need some way to choose who to interview, or no one would ever get on a plane. This is the seductive appeal of programmes such as SPOT and FAST.

But, to Honts, the decade since the 11 September attacks has been one of lost opportunity. Calling SPOT an "abject failure", he says that the government would have done better to invest first in basic science, experimentally establishing how people with malintent think and respond during screenings. That work, in turn, could have laid a more solid foundation for effective detection methods.

Granted, Honts says, that measured approach would have been slow, but it would have been a better investment than rushing to build hardware first, or implementing programmes before they have been tested. "We spent all this time, and all this money," he says, "and nothing has been accomplished." ∎

Sharon Weinberger is a freelance reporter based in Washington DC.

1. Hontz, C. R., Hartwig, M., Kleinman, S. M. & Meissner, C. A. *Credibility Assessment at Portals, Portals Committee Report* (2009).
2. Ekman, P. & O'Sullivan, M. *Am. Psychol.* **46,** 913–920 (1991).
3. Ekman, P., O'Sullivan, M. & Frank, M. G. *Psychol. Sci.* **10,** 263–266 (1999).
4. Bond, C. *Appl. Cognit. Psychol.* **22,** 1298–1300 (2008).
5. Aikins, D. E., Martin, D. J. & Morgan, C. A. III. *Psychophysiology* doi:10.1111/j.1469-8986.2010.00976.x (2010).
6. Leal, S., Vrij, A., Mann, S. & Fisher, R. P. *Acta Psychol.* doi:10.1016/j.actpsy.2010.03.005 (2010).