Adrià Fruitós (Special to ProPublica)

The FBI Says Its Photo Analysis Is Scientific Evidence. Scientists Disagree.

The bureau's image unit has linked defendants to crime photographs for decades using unproven techniques and baseless statistics. Studies have begun to raise doubts about the unit's methods.

by Ryan Gabrielson, Jan. 17, 5 a.m. EST

At the FBI Laboratory in Quantico, Virginia, a team of about a half-dozen technicians analyzes pictures down to their pixels, trying to determine if the faces, hands, clothes or cars of suspects match images collected by investigators from cameras at crime scenes.

The unit specializes in visual evidence and facial identification, and its examiners can aid investigations by making images sharper, revealing key details in a crime or ruling out potential suspects.

But the work of image examiners has never had a strong scientific foundation, and the FBI's endorsement of the unit's findings as trial evidence troubles many experts and raises anew questions about the role of the FBI Laboratory as a standard-setter in forensic science.

FBI examiners have tied defendants to crime pictures in thousands of cases over the past half-century using unproven techniques, at times giving jurors baseless statistics to say the risk of error was vanishingly

small. Much of the legal foundation for the unit's work is rooted in a 22-year-old comparison of bluejeans. Studies on several photo comparison techniques, conducted over the last decade by the FBI and outside scientists, have found they are not reliable.
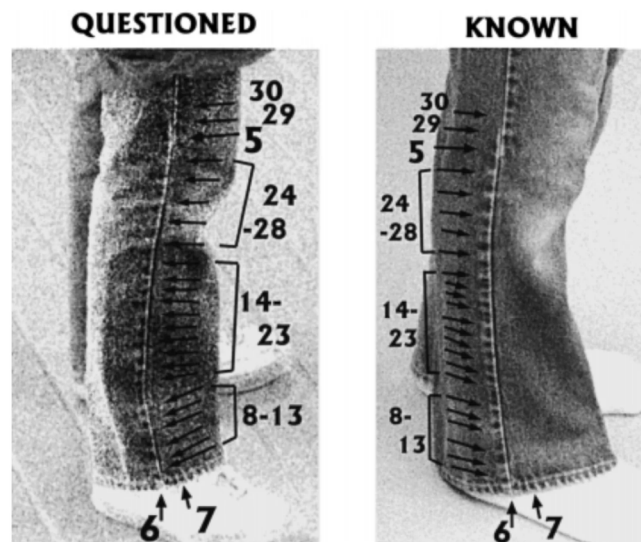

FIG. 7—Side-by-side comparison of Spokane bank robber's right leg ("Questioned") and right leg of model wearing blue jeans recovered from suspect's home ("Known"). Numbers refer to features described in text.

A side-by-side photo comparison of bluejeans from a 1996 robbery and bombing case published with an article on the bureau's clothes identification method. (Journal of Forensic Sciences)

Since those studies were published, there's no indication that lab officials have checked past casework for errors or inaccurate testimony. Image examiners continue to use disputed methods in an array of cases to bolster prosecutions against people accused of robberies, murder, sex crimes and terrorism.

The work of image examiners is a type of pattern analysis, a category of forensic science that has repeatedly led to misidentifications at the FBI and other crime laboratories. Before the discovery of DNA identification methods in the 1980s, most of the bureau's lab worked in pattern matching, which involves comparing features from items of evidence to the suspect's body and belongings.

Examiners had long testified in court that they could determine what fingertip left a print, what gun fired a bullet, which scalp grew a hair "to the exclusion of all others." Research and exonerations by DNA analysis have repeatedly disproved these claims, and the U.S. Department of Justice no longer allows technicians and scientists from the FBI and other agencies to make such unequivocal statements, according to new testimony guidelines *<https://www.documentcloud.org/documents/5684666-Final-Hair-Ultr-07242018-for-Posting.html#document/p3/a475152>* released last year.

Though image examiners rely on similarly flawed methods, they have continued to testify to and defend their exactitude, according to a review of court records and examiners' written reports and published articles.

ProPublica asked leading statisticians and forensic science experts to review methods image examiners have detailed in court transcripts, published articles and presentations. The experts identified numerous instances of examiners overstating the techniques' scientific precision and said some of their assertions defy logic.

The FBI declined repeated requests for interviews with members of the image group, which is formally known as the Forensic Audio, Video and Image Analysis Unit.

ProPublica provided the bureau written questions in September and followed up in November with a summary of our reporting on the bureau's photo comparison practices. The FBI provided a brief prepared response last month that said the image unit's techniques differ from those discredited in recent studies. It said image examiners have never relied on those methods "because they have been demonstrated to be unreliable."

But the unit's articles
<https://www.documentcloud.org/documents/5684669-FBI-Forensic-Identification-Public.html> and presentations
<https://www.documentcloud.org/documents/5684676-87378239-VorderBruegge-Face.html> on photo comparison
<https://www.documentcloud.org/documents/5684686-Fbi-Photocompare-Outline-2007.html> show its practices mirror those used in the studies.

The bureau did not address examiners' inaccurate testimony and other questionable practices.

Judge Jed Rakoff of the United States District Court in Manhattan, a former member of the National Commission on Forensic Science, said the weakest pattern analysis fields rely more on examiner intuition than science. Their conclusions are, basically, "my hunch is that X is a match for Y," he said. "Only they don't say hunch."

Rakoff said that image analysis hadn't come before him in court and wasn't taken up by the commission but said that investigators, prosecutors and judges should make sure evidence is reliable before using it.

Scandals involving other areas of forensic science have shown the danger of waiting for injustices to become public to compel reform, Rakoff said.

"How many cases of innocent people being wrongly convicted have to occur before people realize that there's a very broad spectrum of forensic science?" Rakoff asked. "Some of it is very good, like DNA. Some of it is pretty good, like fingerprinting. And some of it is not good at all."

Details on FBI caseloads and testimony are not readily available to the public. As such, there is no way to determine exactly how often image

examiners testify and when their photo comparisons serve as central evidence in prosecutions. In court, examiners have said they analyze photos in hundreds of cases a year, according to trial transcripts.

To try to identify some of those cases, ProPublica searched court databases and found more than two dozen criminal cases since 2000 in which documents mentioned the FBI's image examiners, nearly all cases that were appealed and thus had a substantial written record. Few criminal convictions, though, make it to an appeal.

None of the appealed cases led to judges reversing convictions, nor has evidence emerged to show that the defendants were innocent. Still, flaws in forensic science techniques often emerge decades after they've been allowed by judges and been used to secure convictions.

The problems with the FBI's photo comparison work plague other subjective types of forensic science, such as fingerprint analysis, microscopic hair fiber examination and handwriting analysis, said Itiel Dror <https://www.ucl.ac.uk/~ucjtidr/>, a neuroscientist who trains U.S. law enforcement on cognitive bias in crime laboratories. Dror is a researcher at University College London, frequently teaching at agencies like the FBI and New York Police Department on ways to minimize personal beliefs from influencing casework.

Even DNA analysis can be swayed by bias <https://www.documentcloud.org/documents/5684672-Dror-SJ-Subjectivity-and-Bias-in-DNA-Mixture.html>, Dror said. But pattern-matching fields like image analysis are especially vulnerable. Image examiners' lab work is, generally, only seeing if evidence from a suspect "matches" that from a crime scene.

"Many of them are more concerned by what the court accepts as science rather than being motivated by science itself," Dror said.

A Plaid Shirt, and Staggering Odds

The image unit has characterized its lab work as akin to that of biologists and chemists. "Just as DNA examiners can point to repeating base pair matches to justify an identification, image examiners must be able to point to actual physical features on a face or body to justify their conclusions in court," an FBI publication from 2008 <https://www.documentcloud.org/documents/5684669-FBI-Forensic-Identification-Public.html#document/p1/a475153> read.

A bank robbery trial 16 years ago was a watershed for such testimony. Prosecutors charged an ex-convict with robbing a string of banks across South Florida over two years. Richard Vorder Bruegge, an FBI image examiner, told jurors that the button-down plaid shirt found in the defendant's house was the exact shirt on the robber in black-and-white surveillance pictures. The examiner said he matched lines in the shirt patterns at eight points along the seams.

The prosecutor asked Vorder Bruegge what were "the odds in which two shirts would be randomly manufactured by the company, having all those eight points of identification lining up exactly the same?"

Only 1 in 650 billion shirts would randomly match so precisely, Vorder Bruegge said <https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p51/a475154>, "give or take a few billion."
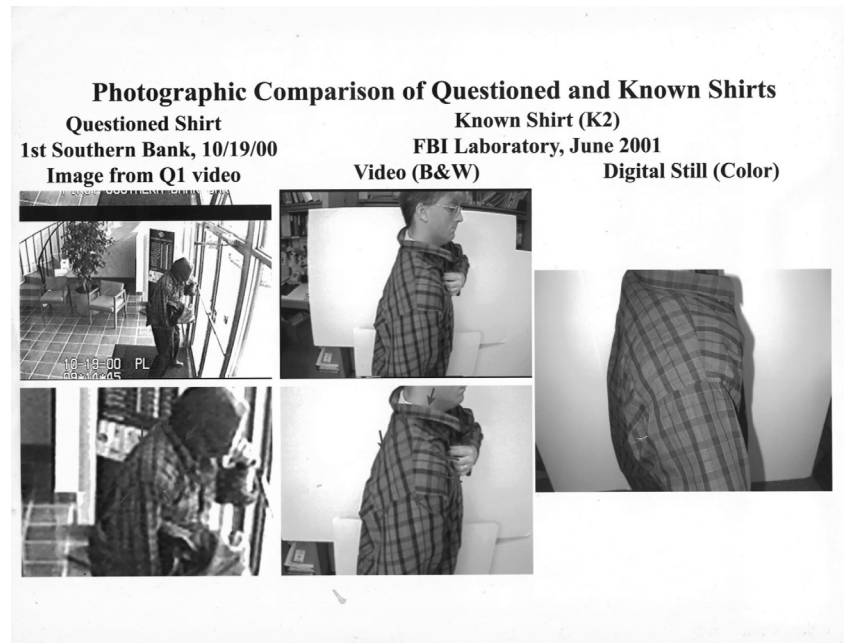
An exhibit showing a 2001 photo comparison of a bank robber's shirt in surveillance video to a shirt seized from a defendant's home and modeled by FBI Lab examiner Richard Vorder Bruegge. The red arrows point to shirt features that allegedly match. (FBI Forensic Audio, Video and Image Analysis Unit, via Wilbert McKreith)

Prosecutors had presented jurors with days of circumstantial evidence against the ex-convict, Wilbert McKreith, before Vorder Bruegge took the stand. Witnesses had seen a burgundy-colored sedan similar to McKreith's Mercedes-Benz outside a majority of the robberies. And McKreith made cash purchases totaling about $10,000 during the period the robberies occurred, even though he had no steady income. (He said he borrowed money from his parents.)

But when the jury convicted McKrieth, who is serving a 92-year prison sentence <https://www.documentcloud.org/documents/5684673-Mckreith-Sentence-2003.html#document/p1/a475155>, Vorder Bruegge's photo comparison and statistics were the only evidence <https://www.documentcloud.org/documents/5684674-US-V-McKreith-Trial-Day-5.html#document/p24/a475157> that had directly connected the defendant to the crime spree.

The statistics were also preposterous, seven statisticians and independent forensic scientists told ProPublica.

The features Vorder Bruegge matched might be common in plaid shirts, making them of little value for identifying the garment, said Karen Kafadar *<https://forensicstats.org/team/karen-kafadar/>*, chair of the statistics department at the University of Virginia. No one has studied the alignment of lines on men's button-down shirts. There is no database of shirt features allowing Vorder Bruegge to calculate the probability of a random match, a statistic used to explain results from DNA typing.

Kafadar has worked in forensic science validation for the past two decades, contributing to a groundbreaking study of the FBI's bullet lead analysis. She said Vorder Bruegge's statements are brazen.

"Somehow they feel perfectly entitled to make outrageous statements," said Kafadar, who said the 1-in-650-billion claim "makes about as much sense as the statement two plus two equals five."

Photographic Comparison of Questioned and Known Shirts

Questioned Shirt
Bank United, 12/5/00
Image from Q1 video

Known Shirt (K2)
FBI Laboratory
June 2001

In this photo comparison exhibit, Vorder Bruegge placed white arrows pointing to lines he contended helped identify the defendant's shirt as the one worn by the robber in a December 2000 bank heist. (FBI Forensic Audio, Video and Image Analysis Unit, via Wilbert McKreith)

Research has bolstered some of the image unit's practices. Last year, a federal study determined that professional image examiners matched faces more accurately *<https://www.documentcloud.org/documents/5684693-Nist-Blackblox-Pnas.html>* than untrained students — providing the first scientific basis for a photograph comparison field.

However, in the past six years, FBI examiners participated in preliminary tests on techniques for identifying faces and hands in pictures with skin features — freckles, folds and creases, moles and blemishes. In both, participants couldn't consistently mark the features to use in an analysis. They marked a certain number of creases or freckles on a face or hand the first time and came up with very different counts the next.

Those studies found alarming inconsistency. If examiners cannot mark the same features each time they use a technique, "then you can't rely on the result, I think that's what any statistician would say," said David Kaye <https://pennstatelaw.psu.edu/faculty/kaye>, a Penn State University law professor and expert on DNA analysis. "It's not a reliable measure."

The FBI response to ProPublica said the image unit's own methods differ from those in the studies. But the unit's published descriptions of its practices <https://www.documentcloud.org/documents/5684676-87378239-VorderBruegge-Face.html#document/p58/a475159> show they are effectively the same as the ones tested by researchers.

Image examiners testified about conclusions based on these methods as recently as last year. The lab has not conducted, or has not published, similar research on its techniques for matching clothes or cars in pictures <https://www.documentcloud.org/documents/5684664-Research-Needs-Assessment-VITAL-Vehicle.html#document/p2/a475151> .

In 2014, an FBI face comparison contributed to a wrongful arrest in Colorado, detailed in an article by The Intercept <https://theintercept.com/2016/10/13/how-a-facial-recognition-mismatch-can-ruin-your-life/>. But image analysis has otherwise drawn little scrutiny.

Such deficiencies rarely matter in court. Few defense lawyers receive training in science or statistics, leaving them ill-suited to dispute expert witnesses. Examiners from Quantico seem "virtually unchallengeable" on the stand, said Lara Bazelon, a former federal public defender who is now a University of San Francisco law professor.

"I think everyone looking back has regrets about things that they've done as a lawyer," Bazelon said, "but one of mine certainly is accepting a lot of the science that I got in discovery as, 'Well, it came from the FBI Lab and it sounds really sophisticated, so it's probably true.'"

For the Nation's Crime Lab, a Reckoning

The FBI opened the laboratory in 1932, and for the next 60 years its forensic science was more revered than scrutinized.

Then, in August 1995, lawyers for a defendant on trial in the bombing of the World Trade Center called Frederic Whitehurst, a chemist on the bureau's explosives unit, to testify. Whitehurst told the court his lab colleagues had produced inaccurate reports in the case.

He had complained within the lab for years about unqualified explosives examiners and shoddy scientific practices. The FBI mostly dismissed the concerns and, Whitehurst said, reassigned him to a different unit as retaliation. So he went public — on the stand and to the press.

The Justice Department's Office of the Inspector General was already investigating Whitehurst's allegations. Its final report <https://oig.justice.gov/special/9704a/index.htm>, released in 1997,

confirmed the explosives unit had "significant instances of testimonial errors, substandard analytical work, and deficient practices" in several cases, including the World Trade Center and Oklahoma City bombings.

Officials at the bureau had overlooked the explosives unit's bad practices and didn't move urgently to fix them, Bill Esposito, then FBI deputy director, said at the time

> " We found, however, significant instances of testimonial errors, substandard analytical work, and deficient practices.
>
> *U.S. Department of Justice, Office of the Inspector General's Special Report, 1997*

<https://fas.org/irp/agency/doj/oig/fbilab1/espo.htm>. "While the issues raised by the Inspector General concern only a small part of the total volume of work done annually in the lab, we recognize that even one problem is too many."

The lab already knew about a second problem.

As the explosives unit became a scandal, the Justice Department began reviewing the top FBI hair and fiber examiner's work on hundreds of cases.

The in-house review looked at examiner Michael P. Malone's lab work and sworn statements in more than 250 cases. It found Malone routinely misrepresented his results as valid and his error rate as less than 1 percent. Justice Department officials did not make the finding public, nor did it notify lawyers for the defendants in those cases, or scrutinize the rest of the hair unit, reporting by The Washington Post <https://www.washingtonpost.com/local/crime/convicted-defendants-left-uninformed-of-forensic-flaws-found-by-justice-dept/2012/04/16/gIQAWTcgMT_story.html?utm_term=.1e6f982faa15> revealed in 2012.

Advances in DNA analysis technology were rattling many forensic science fields, revealing wrongful convictions won with other crime-lab evidence. Microscopic hair comparisons were particularly vulnerable to debunking because follicles contain genetic material. For decades, examiners told jurors that crime scene hairs came from defendants. DNA analysis later proved the hairs did not in dozens of cases. (The FBI replaced microscopic hair comparisons with DNA in 1999.)

Prompted by the Post's investigation, the Justice Department finished an expansive review <https://www.fbi.gov/services/laboratory/scientific-analysis/fbidoj-microscopic-hair-comparison-analysis-review> of hair comparison testimony. Hair examiners matched defendants to follicles in 268 trials; all but 11 contained scientific error. They were more conservative in their written lab reports, about half of which included a misstatement. Like other forensic science reckonings, the public disclosure came years after the FBI stopped relying on the method.

Another unit at the FBI Lab had for decades matched bullets by their chemical compositions.

FBI chemists asserted the mix of elements in a round could determine whether its lead matched ammunition seized from defendant's cars and homes. In court, they said crime scene rounds were "indistinguishable" from the suspects' bullets, even suggesting they came from the same box. The bureau had no science to back its claims.

Facing court challenges, the FBI in 2002 asked the National Academies of Sciences, Engineering and Medicine to study the methods and value of bullet lead analysis.

The report by researchers <https://www.nap.edu/read/10924/chapter/1> in 2004 said the examiners' testimony went further than the chemical analysis allowed. "The available data do not support any statement that a crime bullet came from a particular box of ammunition," the academies' report <https://www.documentcloud.org/documents/5684677-Cbla-Natresearchcouncil.html#document/p20/a475158> said.

Further, one bullet could match anywhere from 12,000 to 35 million other bullets. FBI officials discontinued lead analysis a year later.

Also in 2004, fingerprint examiners wrongly matched <https://oig.justice.gov/special/s0601/final.pdf> a print from a train bombing in Spain to a lawyer and Muslim convert in Portland, Oregon. Agents arrested and detained the lawyer for more than two weeks, without criminal charges, before Spanish law enforcement disproved the FBI's conclusion.

Following each scandal, the bureau moved to shutter the discredited unit or correct the disputed method. It did not comprehensively search past casework for convictions based on the lab's inaccurate evidence, nor did it evaluate whether other units had the same bad practices — unproven techniques, fabricated error rates, misleading testimony.

"The FBI Lab is a fixer," Whitehurst said in an interview last year. Examiners have many incentives to find evidence that helps a conviction, he said.

In 2009, the National Academies of Sciences published a wide-ranging evaluation of the forensic sciences and their deficiencies. It recommended crime labs be moved out of the police and prosecuting agencies that have always run them. To date, Houston and Connecticut are the only jurisdictions that have made their crime labs independent of the police. The Justice Department never publicly considered separating Quantico from the FBI.

A 2016 report by former President Barack Obama's council of science and technology advisers highlighted the lack of validation <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_scie in several pattern evidence fields. It also called on the FBI to dramatically increase spending on studies to prove its methods. U.S. Department of

Justice officials rejected most of the advisers' conclusions <https://www.fbi.gov/file-repository/fbi-pcast-response.pdf/view>. Federal law enforcement has doubled down on unproven forensic science.

In 2017, then-Attorney General Jeff Sessions closed the National Commission on Forensic Science

> **" The report makes broad, unsupported assertions regarding science and forensic science practice.**
>
> *The FBI response to a 2016 report by a presidential advisory panel criticizing pattern evidence. Sept. 20, 2016.*

<https://www.justice.gov/opa/pr/attorney-general-jeff-sessions-announces-new-initiatives-advance-forensic-science-and-help>, ending an effort to set standards for crime laboratory practices. The department also stopped its internal review of testimony <https://www.washingtonpost.com/local/public-safety/sessions-orders-justice-dept-to-end-forensic-science-commission-suspend-review-policy/2017/04/10/2dada0ca-1c96-11e7-9887-1a5314b56a08_story.html?utm_term=.3217cd357202> from FBI pattern evidence units.

In a law review article last year, two high-ranking FBI Lab scientists dismissed <https://www.documentcloud.org/documents/5684681-Isenberg-Oien-DOJ-39.html> validation concerns as uninformed. They wrote that, already, "every forensic discipline practiced in an accredited forensic laboratory must demonstrate that it is reliable, accurate, and fit for its intended use."

Quantico is, indeed, accredited. But the lab has never proven photo analysis is reliable. It has increasingly done the opposite.

Science and the Supreme Court

From its beginnings, photo comparison has been a craft and FBI image examiners more like tradespeople than scientists. Methods are taught through apprenticeships, with new examiners doing casework alongside lab veterans.

After Congress passed a law in 1968 requiring banks to have security equipment, most banks installed surveillance cameras. Meanwhile, Eastman Kodak sold the public millions of pocket-size cameras and amateur photographers took billions of exposures of life and, occasionally, of crimes.

Pictures flooded the bureau as evidence. The lab formed a team called the Special Photographic Unit to find information in images and manage the bureau's inventory of 35 mm cameras. No scientific background or advanced degrees were required.

The analysis was rarely straightforward, said Gerald Richards, who led the photo unit in the 1980s and early 1990s and is now retired. Photographs were fuzzy and poorly lit, especially those from bank cameras. Robbers often wore masks. When a criminal's face was obscured, they looked at the ears, shirts, pants and shoes.

Fingerprint examiners focus only on the swirls and deltas on human fingertips. Hair fiber examiners only analyzed hairs and fibers.

But image examiners created a tapestry of techniques that cross into photography, physics, clothes manufacturing, dermatology, auto body design, human aging and statistics. Still, the unit requires examiners to study photography and little else before working on criminal cases. There weren't even formal courses on photo comparison until 2005, court records show.

Judges long accepted examiners' testimony as expert opinion without much debate. Agents were experts because they worked at the FBI Lab.

Then, in June 1993, the Supreme Court transformed the law around scientific evidence. The court's ruling in Daubert v. Merrell Dow Pharmaceuticals Inc. *<https://supreme.justia.com/cases/federal/us/509/579/>* said federal judges need to assess "whether the reasoning or methodology underlying the testimony is scientifically valid" before allowing it at trial.

Methods should be tested, the opinion said, and results should be based on reliable data that includes error rates. None of the pattern evidence fields met that standard. The Daubert decision posed an existential threat to many forensic sciences.

A month later, the image unit dodged a legal mine set by Daubert. The 9th U.S. Circuit Court of Appeals heard arguments on a bank robbery conviction in Southern California. A jury had convicted James D'Ambrosio based in large part on an FBI image analysis of denim jeans in surveillance pictures. A scientist for the defense *<https://www.documentcloud.org/documents/4638635-US-v-DAmbrosio.html#document/p2/a445385>* testified that clothing comparison was unproven. The appellate court upheld D'Ambrosio's conviction without weighing the scientific merit.

Clothes comparison escaped without damage. But all of the unit's methods seemed vulnerable to challenge. The image unit was filled with former field agents and lab technicians, few of whom held advanced degrees. None had a background in research or academic publishing.

That changed in 1995 when the FBI hired Vorder Bruegge *<https://www.documentcloud.org/documents/4466937-Vorderbruegge-Resume-2005.html#document/p1/a475160>*, the scientist whose testimony about plaid shirts would help prosecutors obtain a conviction in the string of Florida bank robberies.

Before the FBI, Vorder Bruegge, then 31, had spent the previous four years working for a NASA contractor and vying for a spot in the space program

> **❝ However, he said that there was insufficient scientific research done in the field of clothing comparison through photographs to state with any confidence that the jeans were the same.**
>
> *In 1993, a defense expert disputed the scientific validity of an FBI bluejeans comparison, raising questions of whether the image analysis should be admissible. (9th U.S. Circuit Court of Appeals Unpublished Deposition 9 F.3d 1554)*

<https://www.documentcloud.org/documents/4494560-usvFRABIZIO-Daubert.html#document/p13/a427502>. He earned a doctorate in geology from Brown University, where he had studied Venus' mountain belts *<https://www.documentcloud.org/documents/4624377-Vb-Dissertation-06.html#document/p1/a475161>* and had written for science journals.

An Explosion and a "Barcode" for Bluejeans

In the Daubert opinion, the Supreme Court listed validation testing as the best way to meet the evidence standard. Those studies can be complicated to organize and are risky. What if the results disprove what examiners have said under oath for decades?

The next option was peer review, a term indicating the methods had been vetted by outside experts, then published in a science or academic journal. Vorder Bruegge was soon at work on an article that would put denim jean identification, the technique already challenged in court, into the scientific literature.

On April 1, 1996, a bomb had exploded in a lobby of The Spokesman-Review in Spokane, Washington. As police rushed to the newspaper's office, three men robbed a nearby bank and detonated another bomb on their way out.

A similar attack followed three months later at a Planned Parenthood clinic in Spokane and the same bank branch. The bombs caused building damage but no injuries. Surveillance video showed three men in ski masks, heavy jackets and denim jeans.

Agents arrested the members of a right-wing militia group and after searching their homes, seized 27 pairs of jeans. Back at the lab, Vorder Bruegge compared the pants against still images from bank video. He concluded that a pair of the defendants' jeans were identical to those worn

by one of the attackers in the first robbery, and therefore must be the same pants.

Shortly after the militia members received life sentences in prison, Vorder Bruegge submitted an article to the Journal of Forensic Sciences titled, "Photographic Identification of Denim Trousers from Bank Surveillance Film <https://www.astm.org/DIGITAL_LIBRARY/JOURNALS/FORENSIC/PAGES/JFS14519J.htm?citdrop=apa>." The article implied his method of jeans identification was a novel technique, though the photograph unit had long used it.

Vorder Bruegge said each light or dark patch of denim was a unique characteristic and, taken together, they formed a "barcode." <https://www.documentcloud.org/documents/4404918-Vorderbreugge-Jeansidentification.html#document/p3/a475163>

Side seams serve as barcode for identification. (p. 3) <https://www.documentcloud.org/documents/4404918-Vorderbreugge-Jeansidentification.html#document/p3/a475163>

View the entire document with DocumentCloud <https://www.documentcloud.org/documents/4404918-Vorderbreugge-Jeansidentification.html#document/p3/a475163>

Vorder Bruegge's article, "Photographic Identification of Denim Trousers from Bank Surveillance Film," asserted that "barcode" patterns along seams on jeans are unique. (Journal of Forensic Sciences)

The seized pants were J.C. Penney "plain pocket" jeans <https://www.documentcloud.org/documents/4404918-Vorderbreugge-Jeansidentification.html#document/p6/a475162>, which the department store chain marketed as nearly indistinguishable <https://www.youtube.com/watch?v=Fl3s3HZIPcJ4&feature=youtu.be> from more expensive Levi's 501 jeans.

Designs might be similar, but wear marks reveal to FBI examiners the jeans' true identity, Vorder Bruegge argued.

At , he acknowledged the method was not validated. That was of little concern, he wrote, because "the presence of such a large number of significant characteristics in a known pair of blue jeans precludes the possibility (or probability) of their having occurred by mere coincidence."

In requests for responses from the FBI, ProPublica repeatedly sought an interview with Vorder Bruegge; the FBI declined.

With the article, Vorder Bruegge advanced the legal argument for image analysis further in three years than the FBI Lab had the previous three decades. It helped an array of methods <https://www.swgit.org/legal> meet the Daubert standard and become admissible scientific evidence in criminal trials.

Leading forensic scientists, statisticians and clothes manufacturing experts reviewed Vorder Bruegge's article at ProPublica's request. They said the FBI examiner's central claims were misleading or wrong.

He wrote that manufacturing defects like dropped stitches, where a stitch is missing, are identifying features — the equivalent of a facial scar.

Not at all, said Alicia Carriquiry <https://forensicstats.org/profile/?smid=1475>, director of the Center for Statistics and Applications in Forensic Evidence and an Iowa State University professor. Sewing machines can drop stitches in a consistent manner, embedding the same set of stitches in garment after garment.

"This could be that the same sewing machine in China is producing a drop stitch in the same position in every last pair of jeans until they change that needle," Carriquiry said. Thousands of pairs of jeans would have the same feature.

The barcode pattern is unique because the stitching varies between pairs, Vorder Bruegge wrote.

But jean manufacturing has been standardized across the industry for a long time, said Charles Jebara, chief executive of Alpha Garment, which sells jeans under Nicole Miller and other labels. The number of stitches per inch along a seam is much the same from one factory floor to another. "They're using the same kinds of machines, the same general processes to get that operation done," Jebara said.

Denim in various pairs of jeans is so similar that the FBI's hair and fiber unit long ago deemed it useless as evidence. "Because of the commonality of blue denim cotton fibers, we don't even bother to compare them in the FBI Laboratory," an examiner testified <https://www.documentcloud.org/documents/5026761-usvYee-denim-fiber.html#document/p119/a475164> in a 1991 murder trial.

<https://www.documentcloud.org/documents/5026761-usvYee-denim-fiber.html#document/p119/a475164>

Case: 3:89-cr-00720-JZ Doc #: 494 Filed: 02/19/91 119 of 170. PageID #: 12322

1513

View the entire document with DocumentCloud <https://www.documentcloud.org/documents/5026761-usvYee-denim-fiber.html#document/p119/a475164>

```
 1   A      Yes.
 2   Q      What about on the fibers, let's say, in Levi's blue jeans?
 3   A      I agree with you.
 4   Q      Blue cotton.  And so if I show you a pair of what we call
 5          a known pair of blue cotton pants, for example, pair of blue
 6          cotton Levi's and I gave you sweepings from the homes of
 7          everybody in this room, it would not -- it would be likely,
 8          would it not or at least very possible, that, in fact, you would
 9          find these blue cotton threads in almost everybody's home?
10   A      They are very common, yes.
11   Q      And if I ask you to make a report on that, you would
12          testify, would you not, or you would opine that indeed it is
13          consistent that the person wearing those pair of pants had been
14          in the room -- in this courtroom?
15   A      Because of the commonality of blue cotton fibers, we
16          don't even bother to compare them in the FBI laboratory.
17   Q      But if you were asked to do so, the answer would be yes,
18          they are consistent, it would be silly, after a fashion, right?
19   A      Yes.
20   Q      But indeed you would opine that indeed it's consistent
21          that this person wearing this one pair of pants had been in
22          everybody else's home?
23   A      No.
24   Q      Although indeed it would be far more consistent that these
25          fibers originated independently and in fact that's why you don't
```

An FBI fiber examiner testified in a 1991 murder trial that denim fibers are useless as evidence. (U.S. District Court for the Northern District of Ohio, Western Division)

Suzanne Bell <https://forensics.wvu.edu/faculty-and-staff/faculty/suzanne-bell>, head of the forensic science department at West Virginia University, read Vorder Bruegge's article when it was first published. Its flaws mirror those of many fields: examiners making statements they cannot prove, Bell wrote in response to ProPublica's questions.

"The problems usually arise in over-selling the value and implying probabilistic information when there really is none," she wrote. "You can see that in the article."

Jeans comparisons could help ongoing investigations, but they aren't conclusive evidence. "It wouldn't stand scrutiny today," Bell wrote.

A Bank Robbery Case Puts Clothing Analysis on Trial

The FBI Lab put its new scientific literature to use in Florida.

In 2002, prosecutors charged Wilbert McKreith, an ex-convict and entrepreneur, with robbing eight banks along the Fort Lauderdale coastline. The evidence gathered against McKreith wasn't overwhelming.

But agents had seized a plaid button-down shirt from McKreith's house and sent it to the image unit at Quantico for analysis. Vorder Bruegge got the case. <https://www.documentcloud.org/documents/5026761-usvYee-denim-fiber.html#document/p119/a475164>

In the lab, he put on McKreith's shirt <https://www.documentcloud.org/documents/5684684-Mckreith-Fbilabphoto-07.html#document/p1/a475165> and stood in poses similar to the robber in surveillance pictures while another examiner took photographs. Vorder Bruegge compared pictures of himself to those of the robber, focused exclusively on how parts of the shirt lined up along the seams.

On many mass-produced shirts, the lines on one section don't align with those of other sections, causing the patterns to clash where they're stitched together. FBI image examiners routinely testified those clashing patterns were "individual characteristics" that can identify a garment.

In the case of U.S. v. McKreith, Vorder Bruegge took the old method one giant leap further by adding statistics. He concluded the defendant's shirt matched the robber's at eight different points, court records show *<https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p51/a475154>*. And then he calculated the probability that a random shirt — not McKreith's — would match as precisely.

Measuring the pixelated bank photo *<https://www.documentcloud.org/documents/5684654-Mckreith-Fbiphoto-Kislak-02.html#document/p1/a475166>*, Vorder Bruegge decided the odds that one feature would match on a random plaid shirt were only 3 percent *<https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p50/a475167>*. If two features matched, the random match probability dropped to one-tenth of a percent.

For all eight features, the chance that a shirt other than McKreith's would match was just 1 in 650 billion, the examiner decided.
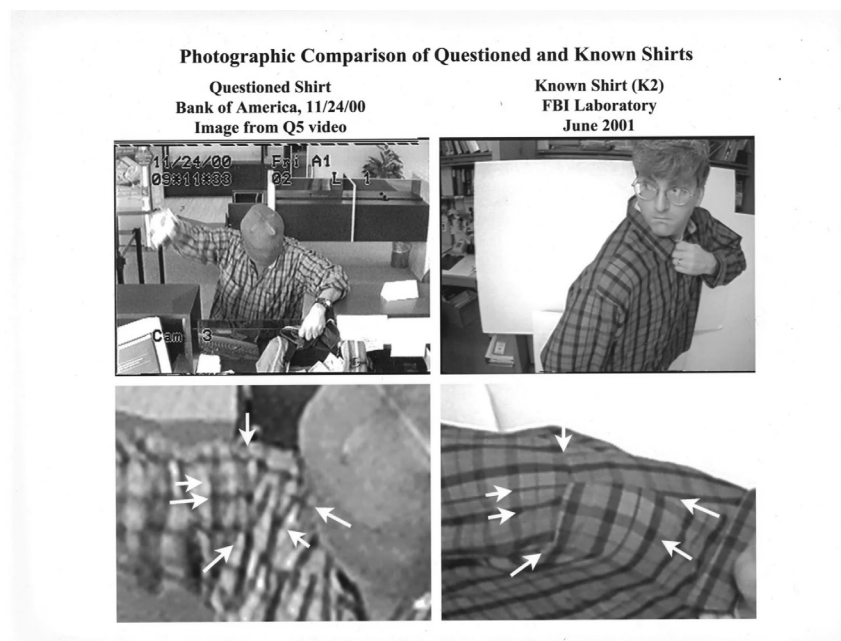
Prosecutors used Vorder Bruegge's testimony in an effort to erase any doubt *<https://www.documentcloud.org/documents/5684674-US-V-McKreith-Trial-Day-5.html#document/p19/a475168>* about McKreith's guilt. In the Fort Lauderdale federal courthouse, the FBI examiner cited his Journal of Forensic Sciences article on jeans comparison to establish his methods were valid.

John Howes, McKreith's defense attorney, asked the court to suppress the image analysis as unscientific. But he didn't see the article before they were in court, and he never read it.

The judge ruled Vorder Bruegge's testimony met the Daubert standard and was admissible. The decision enshrined the FBI unit's techniques and testimony as reliable scientific evidence.

Near the close of McKreith's trial, Roger Stefin, an assistant U.S. attorney, asked Vorder Bruegge what his analysis determined about McKreith's shirt and the robber's shirts in pictures from several banks.

"They're all the same shirt," he said *<https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p58/a475156>*.

Photographic Comparison of Questioned and Known Shirts

Vorder Bruegge matched features on the robber's shoulder, where the shirt fabric was bunched and blurred, in a surveillance image of a November 2000 bank robbery. He'd previously written that precise measurements of clothes in pictures are often unreliable. (FBI Forensic Audio, Video and Image Analysis Unit, via Wilbert McKreith)

In fact, Vorder Bruegge's original analysis did not link <https://www.documentcloud.org/documents/5684685-Mckreith-Fbilabreport-03.html#document/p5/a475170> McKreith's plaid button-down shirt in one of the earliest robberies, of a Commerce Bank branch in May 2000, according to the written lab report. The surveillance images were not detailed enough and "it was not possible to identify" the defendant's shirt "as the shirt worn by the robber to the exclusion of all other shirts."

Vorder Bruegge directly contradicted his report <https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p54/a475171> in court. He explained at length to jurors how he matched shirts in that case, with four large Commerce Bank photo exhibits.

The jury convicted McKreith of seven robberies. Now 60, he is held at a federal penitentiary in California's Central Valley, with 76 years remaining on his sentence. He's exhausted his appeals, most of which attempted to dispute the FBI Lab findings.

The statisticians who reviewed Vorder Bruegge's materials for ProPublica said the examiner's calculations cannot be correct. Vorder Bruegge's statistic — 1 in 650 billion — is simply too astronomical to be true, said Kaye, the Penn State professor. There isn't a database documenting features on plaid-shirt seams like there is for human DNA, making it impossible to determine the likelihood a different shirt would appear to match the robber's shirt.

Many problems in the examiner's testimony went unnoticed, or were simply unknown, during trial. For example, Vorder Bruegge undercut the precision of his calculations when he admitted having rounded down <https://www.documentcloud.org/documents/5684229-US-V-Mckreith-Trial-Day-4.html#document/p50/a5> the shirt measurements used in his calculations because "it makes the math easier."

The jeans article, which Vorder Bruegge cited as proof his methods are accepted science, does not mention any of the techniques he used in the shirt comparison.

Further, Vorder Bruegge wrote in the article that measurements of objects in photographs

> **It would be one in thirty-five times one in thirty-five. But to simplify things and to be conservative, I prefer to use one in thirty. By saying one in thirty, that's — it's giving it a better chance of being the same, but it makes the math easier. Thirty times thirty is nine hundred.**
>
> *Vorder Bruegge's testimony in U.S. v. McKreith. (Transcript from U.S. District Court for the Southern District of Florida)*

<https://www.documentcloud.org/documents/4404918-Vorderbreugge-Jeansidentification.html#document/p2/a475172> are "less accurate when measuring curved objects such as a draped trouser leg," the article said. The photographed shirts in the McKreith case were curved around shoulders and arms.

On the stand, Vorder Bruegge didn't mention that his precise measurements might be inaccurate.

"It may be an honest belief," Kaye said, "though terribly flawed."

Five years after the trial, Vorder Bruegge described his methods <https://www.documentcloud.org/documents/5684686-Fbi-Photocompare-Outline-2007.html> in a presentation to the National Academies of Sciences, including clothes identification and random match probabilities.

Testing the Techniques, and Coming Up Short

Image examiners at the bureau have boasted they can figure out who's who in photographs even under the most difficult circumstances so long as they can see details on the skin. Scars, tattoos and chipped teeth make identifications straightforward.

Examiners contend they can do the same with only common skin marks: freckles, blemishes, wrinkles, creases on the lips.

"By using these traits, effectively the 'texture' of the face, examiners have been able to differentiate between identical twins in images," members of

the unit wrote <https://www.documentcloud.org/documents/5684669-FBI-Forensic-Identification-Public.html#document/p2/a475173> in an FBI publication.

The same principle has been applied to the back of suspects' hands. In some cases, most commonly sexual assaults, the assailant takes pictures of their criminal act and one of their hands stays in the frame. Investigators find the images on a computer hard drive and want to confirm the photographed hand belongs to their suspect.

Military investigators asked the FBI Lab for such an analysis in 2013, in a case involving a U.S. Air Force lawyer accused of raping a child. Christopher Iber, an examiner in the image unit, received the evidence and set about comparing hands.

At trial, Iber "testified that based on similar features between the two hands — such as knuckle creases, hand creases, and blemishes — in his opinion, the hands depicted in the two photographs were the same," an Air Force appellate decision <https://www.documentcloud.org/documents/5684691-usvRichards-Appealopinion.html#document/p75/a475174> states.

Iber did not respond to interview requests.

The military lawyer was convicted at court-martial; he tried to overturn the conviction, in part, by arguing Iber's work was not valid. But the defense didn't challenge the underlying science of hand comparison at trial, and the appeals court dismissed the argument.

Unbeknownst to the courts, the FBI Lab itself was challenging the science behind its skin mark comparisons, somewhat inadvertently.

Vorder Bruegge partnered with Patrick Flynn, a University of Notre Dame computer science professor, on a research project in 2011. They served together on a group writing standards for facial identification by law enforcement.

Facial recognition algorithms match photos primarily by measuring the relative distances between a face's landmarks — specific points on the eyes, nose, brow and so on. Flynn believes adding skin marks to the formulas can help their accuracy. The FBI Lab had already been using those features in image analysis, so Vorder Bruegge lent his experience.

Photos of identical twins were ideal for testing the idea, Flynn said, as their facial landmarks are exactly the same but their freckles and creases were believed to be different. The algorithm would try to locate skin marks, but he had graduate students mark them, too, just as examiners do.

An early finding disputed the FBI's contention that each identical twin had his or her own unique features. Researchers documented that twins share freckles <https://www.documentcloud.org/documents/4436006-Facial-Marks-Identical-Twins.html#document/p6/a475177> much the way they share all other genetic traits.

The FBI's response to ProPublica said the unit's twin comparisons in casework "dating to the early 1990's demonstrate that these individuals can be easily distinguished from one another based on these patterns, when the marks are visible."

But the study continued, next examining how consistently the computer found skin marks compared with the human participants. The algorithm did badly, but the humans were completely unreliable <https://www.documentcloud.org/documents/5684698-Twin-Facial-Marks-Second-Paper.html#document/p6/a475178>. All the participants came up with different sets of freckles and blemishes. Moreover, participants were asked to locate skin features on the same photos twice, and they came up with different results each time.

The study had troubling implications for the FBI's image unit. If examiners mark different features to analyze each time they look at a picture, their entire technique is likely unreliable. Science demands consistent results.

It does not appear the bureau has undertaken a study on its examiners' performance, even as similar research results have continued to come in.

In 2012, the Defense Forensic Science Center, the U.S. military's crime laboratory, tested hand comparisons. Researchers intended to develop an algorithm that could identify people the way the FBI Lab does. They began with the first step in validation, confirming examiners could consistently locate skin features on the back of hands in pictures.

> " ... individual observers perceive facial marks differently over time and the annotations are inconsistent. ... different observers view facial marks differently, leading to inconsistency.
>
> *Article detailing the results of a study on the use of facial skin features to distinguish between identical twins. Vorder Bruegge was one of the co-authors. ("Analysis of Facial Marks to Distinguish Between Identical Twins", IEEE Transactions on Information Forensics and Security, Vol. 7, No. 5, October 2012)*

The results were unexpectedly poor <https://www.documentcloud.org/documents/4624376-Observerreliability-Handcomp-2014.html#document/p4/a475179>. Professional examiners came up with differing sets of freckles and sunspots each time they reviewed the hand images, and they didn't even seem to use the same method as one another.

Most damning, the trained forensic scientists were no more reliable than students. The military researchers published their results in the Journal of Forensic Sciences in November 2015.

"It's another example of the familiar story," said Simon Cole, a University of California, Irvine, sociology professor and pattern evidence researcher.

"Use in court first, validate second."

That did not dissuade the FBI Lab. A bureau image examiner testified on the results of a thumb comparison *<https://www.news-press.com/story/news/crime/2017/05/16/diana-alvarez-guerrero-torres-federal-court/101759600/>* in a May 2017 child pornography trial.

But Vorder Bruegge had taken notice. Around the time of the trial he selected Derek Boyd, an anthropology graduate student at the University of Tennessee, for a summer internship at Quantico solely to conduct an in-house hand comparison study.
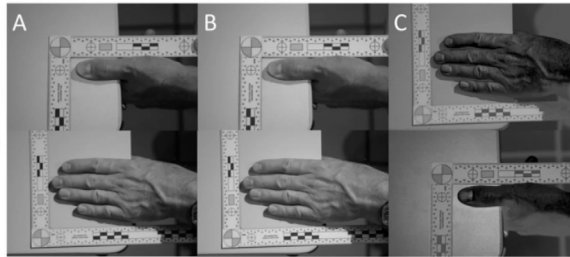

Results

Image 1. Digital images utilized in this study: (A) primary images; (B) images of greater exposure; and (C) primary images with yellow channel isolation in Adobe Photoshop CC.

(1) There is high data dispersion among reported trait counts that is unaccounted for by experience and image exposure

2.00

Table 1. Analysis of variance results for the identification

Vorder Bruegge took pictures of his own left hand, then marked its features as a participant in a study of hand comparison. (Image from poster board presented at the American Academy of Forensic Sciences in February 2018. Courtesy Derek Boyd.)

Three interns and three FBI examiners documented knuckle creases and other skin features on pictures Vorder Bruegge took of his own left hand. Boyd said he expected the results would bolster the hand comparison technique.

Instead, they debunked the method *<https://www.documentcloud.org/documents/4466891-Dorsalhandid-Fbistudy.html#document/p1/a428641>* a second time. Examiners were no better than interns. All were inconsistent and imprecise.

"I was fascinated by how the human eye is still outperforming the algorithm," Boyd said in an interview. "Yet what we found here is the human eyes don't necessarily agree. That's alarming."

Vorder Bruegge and the other examiners had muted reactions when he delivered the study results, Boyd said. "There was just kind of a, 'OK, well, that's good to know,'" he said.

Ryan Gabrielson

Ryan Gabrielson is a reporter for ProPublica covering the U.S. justice system.

✉ ryan.gabrielson@propublica.org

🐦 @ryangabrielson   📱 510-990-6308

🔒 Signal: 917-455-1713