

Firearms forensics is becoming more quantitative

Science over subjectivity could increase juries' confidence in gun identification.

Someone drops a dead body at a hospital. Bullets from the decedent and from a wall where a shooting took place are found to possess matching toolmarks—striations and other markings in the softer bullet material made by imperfections in the gun's barrel. When a suspect's gun is determined to create such

toolmarks, prosecutors argue that the suspect fired those bullets. But if, say, a bullet retrieved from a child has toolmarks that differ from those created by a suspect's gun, "You know it's not the same gun," says Michael Haag, who worked for 25 years in a police crime department and continues his long-time gig as an independent forensics consultant.

For more than a century, examiners have matched guns to crimes by comparing the markings made on fired bullets and spent casings with those test-fired from a known gun. Matching bullets or

the cartridge cases they are held in—which often remain near the shooting site—to guns is a primary source of evidence in hundreds of cases every day across the US.

Determining that a bullet was shot by a particular gun is not easy, says Alicia Carriquiry, a statistics professor and director of the Center for Statistics and Applications in Forensics Evidence at Iowa State University. "To the naked eye, the toolmarks look like cat scratches. There's a lot of subjectivity. The question is, Can you do something more scientific?"

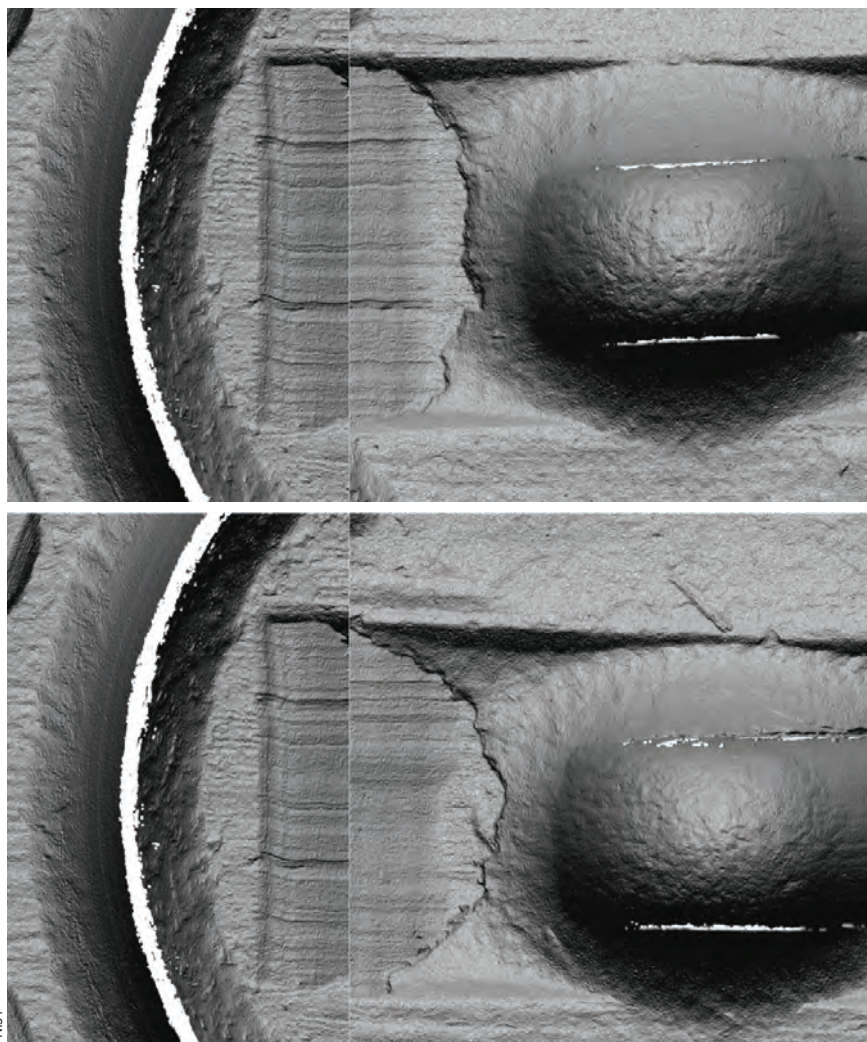
Carriquiry and her colleagues are working to improve toolmark analysis. She calls the results "promising" and expects to see "some changes in the way evidence is evaluated in the next few years."

Meanwhile, scientists at NIST, the Federal Bureau of Investigation, and the Netherlands Forensic Institute are creating a database to help gauge the likelihood that a casing or bullet was fired by a given gun. Xiaoyu Alan Zheng is a mechanical engineer and the NIST lead on the database. He says that the project was "springboarded" by the 2009 National Research Council report *Strengthening Forensic Science in the United States: A Path Forward*. "They wanted more objective comparisons, with scores and numbers, for what is a match or not." The database is expected to go live in three to five years, Zheng says, and will be an ongoing project.

3D advantages

Firearms examiners use traditional microscopy to compare bullets and casings from crime scenes with ones shot from known guns. "There are no real standards on how much similarity there has to be," says Zak Carr, a firearms and toolmark examiner who is now at Cadre Forensics, where he is working on improving and quantifying toolmark identification. When examiners evaluate samples from known sources, the accuracy is high, he adds, noting that research studies show error rates to be around 1%.

Despite the generally high accuracy of examiners' results, says Carr, "it would be nice if you could come up with



A KNOWN MATCH (top), for which two cartridge cases (left and right) were from the same firearm. The bottom image shows two cartridge cases from different firearms. Three-dimensional imaging and improved algorithms are in development to help examiners quantify their assessments on whether a bullet or casing was fired by a specific gun. Such evidence is commonly used in criminal trials.



ANALYSIS FROM THIS CARTRIDGE CASE (left), shot with a 9 mm Luger pistol, is part of a database that the FBI, NIST, and the Netherlands Forensic Institute are working on to help assign a likelihood to whether a bullet or casing was fired by a given gun. The features that would catch an examiner's attention are the lines, the firing pin impression at the center, the drag at 3 o'clock, and the crescent-shaped toolmark running around the firing pin from 7 to 11 o'clock. The center image shows a scan of the cartridge case, which is used to obtain the coordinates of toolmarks. The right-hand image is a map of the depths of the toolmarks made with focus variation scanning; black represents deep spots and yellow represents high spots.

a quantified score for whether toolmarks were from the same or different firearms. That's what we are working toward."

A further impetus for quantifying certainty about evidence, says Carriquiry, is second opinions: Different examiners may reach different conclusions about the same data, or the same examiner looking at the same data again six months later may have a different opinion. And, she says, in recent years "examiners have been taking a beating. Lawyers are waking up to the fact that you can argue against subjective examination. Firearms evidence has been excluded in some cases, which has been a shock."

To set probabilities for source determinations, Carriquiry, scientists at Cadre, and researchers elsewhere are working on analyses of 3D images of bullets, from which "you can look at the depth and position of striations," says Carriquiry. She notes that the algorithms do not always work. And some guns don't mark. "But the crappy guns that most criminals use mark nicely, and the accuracy in identification is high."

Using 3D virtual comparison microscopy has many advantages over traditional examinations, says Carr. "The algorithm can analyze similarities between items in high-volume cases faster. It's huge." The resolution is better, he adds, and the lighting is more consistent. Sharing evidence is also easier. Defense attorneys would like to see criminal evidence, but approval to access and move it can be difficult to obtain. With virtual comparison microscopy, Carr says, "the raw

data can be shared without risk of compromise or loss of evidence."

During his nearly two decades as an examiner, Carr says, he "spent hours hunched over the microscope [and] developed callouses and neck pain. It takes a lot of time to see the details."

Bullet analysis with 3D images "may not make a big difference in results," says Robert Thompson, the senior forensic science research manager at NIST, "but it will augment examiners' subjective opinions with objective data. It will increase confidence for juries."

Not surprisingly, it's often easier to be certain about nonmatches than about matches. And then there are the cases that can't be called. In Carr's experience, some 5–10% of cases have been inconclusive.

Erich Smith is the technical leader for firearms and toolmarks at the FBI Laboratory. In looking through six years of in-house data, he found that just under 14% of cases were reported as inconclusive, with the rest split roughly evenly between identifications and eliminations. Inconclusive is a valid finding, Smith notes. Bullets get damaged, for example. Improved image analysis can make a dent in the rate of inconclusive findings, he and others say.

The algorithms developed in Carriquiry's center are starting to be tested in forensics labs, she says. Some labs already have 3D microscopes, but for others, ponying up a few hundred thousand dollars for one can be a challenge. Introducing 3D microscopes and algorithmic analysis on

a large scale "will take buy-in," she says.

Reference guns and algorithms

NIST, the FBI, and the Netherlands Forensic Institute are building their database from reference guns. The FBI has so far collected scan data for four types of firearms. It has tested about 1000 individual firearms, doing 11 test-fires per gun and using bullets and casings made of different metals. It creates 3D scans of casings and bullets—at about 60 minutes per scan—and uses custom software to score known matches and known nonmatches. "We are building a population with information on the type of gun, manufacturer, surface material, caliber, and more," says Smith. Most important, he adds, the database builds on empirical knowledge about same-source and different-source toolmark comparisons. "In teaching an algorithm to do what humans do, the accuracy of the discipline is increased."

The database project goes back to 2012. When it opens for use, says Smith, examiners will be able to send the 3D images of crime-scene and reference bullets and casings to the FBI team. The team will take those data and assign a score-based likelihood ratio from reference data. Just how the scoring will be done is yet to be determined, says Smith. "That's next on our list."

However the scoring is defined, for a known nonmatch the score is low, and for a known match the dispersion is wider. Importantly, Zheng says, the false-positive rate is low. "You don't want to put the wrong person in jail."

Toni Feder