PUBLISHED

UNITED STATES COURT OF APPEALS
FOR THE FOURTH CIRCUIT

No. 24-4048

UNITED STATES OF AMERICA,

Plaintiff - Appellant,

v.

RON ELFENBEIN,

Defendant - Appellee.

AMERICAN MEDICAL ASSOCIATION; MARYLAND STATE MEDICAL SOCIETY

Amici Supporting Appellee.

Appeal from the United States District Court for the District of Maryland, at Baltimore. James K. Bredar, Senior District Judge. (1:22-cr-00146-JKB-1)

Argued: January 29, 2025 Decided: July 17, 2025

Before AGEE and RICHARDSON, Circuit Judges, and Michael S. NACHMANOFF, United States District Judge for the Eastern District of Virginia, sitting by designation.

Affirmed in part, reversed in part, and remanded by published opinion. Judge Richardson wrote the opinion, in which Judge Agee and Judge Nachmanoff joined.

ARGUED: Jason Daniel Medinger, OFFICE OF THE UNITED STATES ATTORNEY, Baltimore, Maryland, for Appellant. Gregg Lewis Bernstein, ZUCKERMAN SPAEDER LLP, Baltimore, Maryland, for Appellee. **ON BRIEF:** Glenn S. Leon, Chief, Fraud Section, Jeremy R. Sanders, Appellate Counsel, UNITED STATES DEPARTMENT OF JUSTICE, Washington, D.C.; Erek L. Barron, United States Attorney, OFFICE OF THE UNITED STATES ATTORNEY, Baltimore, Maryland, for Appellant. Martin S. Himeles, Jr., ZUCKERMAN SPAEDER LLP, Baltimore, Maryland, for Appellee. Jeff Wurzburg, NORTON ROSE FULBRIGHT US LLP, San Antonio, Texas, for Amici Curiae.

RICHARDSON, Circuit Judge:

According to the United States, two audits, a healthcare-billing expert, four patients, and three employees, Dr. Ron Elfenbein committed healthcare fraud. But according to a different expert, other staff members, and himself, Elfenbein did not. After 11 days of trial, a jury decided that Elfenbein was guilty. But the district court acquitted, reasoning that the jury had too little evidence to convict.

We disagree, so we reverse that decision. But we *do* agree that the case was close—and we find it significant that the most damning evidence came not from the government's witnesses but Elfenbein's. So we affirm the district court's contingent order granting a new trial.

I. Background

A. Elfenbein Runs An Urgent-Care Business

In 2016, Dr. Ron Elfenbein opened an urgent-care clinic in Maryland. Called Drs ERgent Care,[1] the clinic and its satellite locations serve patients in and around Gambrills, a town between Baltimore and Annapolis. During normal times, the clinic's main location was a typical, "full-service urgent care." J.A. 1953. It offered in-person exams, x-rays, lab testing, and "minor in-office procedures," and served about 30 patients daily. J.A. 858.

B. COVID-19 Arrives And Elfenbein's Business Evolves

In the spring of 2020, everything changed. Among many ways the pandemic upended normal life, it made COVID-19 tests all-important—to work, travel, or participate

[1] Today, the clinics operate under a new name: FirstCall Medical Center.

in society. In response to this "overnight demand," Elfenbein tweaked his business model. J.A. 859. The clinics "pivoted away from . . . traditional urgent care services" and toward COVID-19 testing. *Id.* And Elfenbein opened more satellite locations, like one at a fire station in Earleigh Heights, to test more patients. This shift brought a "significant increase" in the number of patients the clinic saw. J.A. 859.

During this time, the clinic mostly operated as a drive-through. Patients who wanted COVID-19 tests could fill out forms in advance, pull into the parking lot, and wait for a nurse to come swab their noses and take their temperatures. Then they would "pull up" under a tent and park next to a television for a virtual appointment, where a provider would appear on the screen and chat with them for a few minutes. J.A. 846. On busy days, the line of cars waiting for tests might wrap around the block. So the clinic moved quickly. One employee described the operation as "moving a herd of cattle through a pass at 60 heads *per minute*!!" J.A. 4497. Or as Elfenbein put it, "[w]e are not there to solve complex medical issues" so "we want them in and out of the tent in under 5 minutes total." J.A. 4487.

Elfenbein's clinic got paid for most of these visits not out of patients' pockets but by insurers like Medicare. Insurance payment requires coordination between insurers (who do not directly observe the provision of medical care) and providers (who do). To simplify and standardize the payment process, providers and insurers classify medical services into general categories and subcategories. Insurers identify these categories with numerical codes. When a provider does medical work, they send the insurer the code that reflects the appropriate category for those services. Then, insurance pays the provider a fixed amount

based on that code. In other words, providers' pay depends on what category a service falls into—not patient- or appointment-specific details. Of course, this system only works if providers use the right codes. To make sure that they do, insurers usually require providers to submit not just codes but documentation that describes the medical services they provided.

To ensure uniformity, many participants in this system use the same coding system. That system comes from an annual American Medical Association guidebook called the *CPT Manual*, for "Current Procedural Terminology." But although the *CPT Manual* lays out the framework, different insurers pay different rates for the same codes. Medicare, for instance, bases its payments on regulations promulgated by a federal agency called Centers for Medicare and Medicaid Services. Along with setting rates, CMS uses regulations to tweak the definitions associated with codes.

What code a provider should use to describe his services thus depends on the interaction between multiple sources. In general, the codes are defined by the latest edition of the *CPT Manual*. Then, the provider should account for any insurer-specific adjustments to the *Manual*'s definitions—like those created by CMS for Medicare. And last, insurers generally require the provider to submit medical documentation showing that the code he used matches the work he did.[2]

[2] For counts one through three, the payor was Medicare. For counts four and five, the payor was CareFirst. Neither party argues that these payors' rules differed in a relevant way.

When an insurer receives this information, it must evaluate the claim and decide whether to pay it. Whether it pays depends, among other things, on whether the service was "medically necessary," whether it was "actually . . . provided . . . as stated on the claim," and whether it is "supported by medical records." J.A. 365–66.

This case arises out of the way Elfenbein's clinic coded five visits. The five named patients visited Elfenbein's clinic between March 5 and May 12, 2021. Each was tested for COVID-19; if they got any further medical treatment, it was typically limited to checking basic vital signs. Some had symptoms, and some did not. But all testified that their visits were short—five or ten minutes apiece.

These visits, all agree, fell into the general category of "evaluation and management" visits. E/M services, under the *CPT Manual*, are divided into two overarching categories. One set of codes applies to evaluation and management for established patients—patients that the provider has seen in the last three years. The second set applies to new patients. Within each set, any given E/M visit falls within one of five levels. A level-one visit is the simplest (and cheapest). A level-five visit is the most complex (and costly). Elfenbein's clinic billed the five visits in question at level four, using code 99204 for four new patients and code 99214 for one existing patient.[3]

This level-four coding generated lots of money for Elfenbein and his clinic. Though many level-four visits took only a few minutes of his clinic's time, Elfenbein charged

[3] The government argued that these visits were representative because, as some of its evidence suggested, Elfenbein apparently instructed his staff to code *all* visits for COVID-19 tests at level four (or five if the patient was symptomatic).

$354.22 for each 99204 visit and $231.50 for each 99214 visit. With nearly 1500 patients coming through on some days, the clinic made millions.

C. The United States Brings Criminal Charges

Where Elfenbein saw opportunity, a federal grand jury saw fraud. When it learned of the clinic's coding practices, the grand jury indicted Elfenbein for five counts of healthcare fraud in violation of 18 U.S.C. § 1347. Each count corresponded to one patient visit for COVID-19 testing conducted during the spring and summer of 2021. As the United States saw things, for each visit, Elfenbein committed fraud in two ways. First, the United States alleged that Elfenbein billed insurers too much for the simple diagnoses he provided. Second, the United States alleged that Elfenbein supported his overbilling by sending insurers medical reports that reflected services his clinic never provided.

D. A Jury Votes To Convict Elfenbein, But The Court Acquits

Elfenbein went to trial. The government relied on an expert witness, Stephen Quindoza, who explained the CPT code system to the jury. Quindoza also opined that level-four codes were generally too high for the quick and easy task of testing someone for COVID-19. But Quindoza did not specifically testify that Elfenbein's coding was improper, and he admitted on cross-examination that he was unfamiliar with the latest, pandemic-era coding rules. The government also called staff from Elfenbein's clinic, many of whom expressed some discomfort with Elfenbein's coding practices. And alongside these witnesses, the government also showed the jury internal emails and patient medical records.

After the government finished its case-in-chief, Elfenbein moved for a judgment of acquittal. The district court denied his motion, concluding that the government had presented enough evidence to convict, and trial continued. In his defense, Elfenbein offered another expert, Michael Miscoe, who provided a fuller explanation of how the CPT coding system works and what level-four codes require. After explaining the system, Miscoe also opined that Elfenbein's coding decisions were correct. Finally, Elfenbein himself testified about the treatments he provided and the codes he used.

The jury returned guilty verdicts on all charges, and Elfenbein again moved for a judgment of acquittal. This time, the district court granted the motion. As it saw things, after the whole trial, "the level 4 codes used to describe the five encounters" may or may not have been false because the codes' definitions were ambiguous. J.A. 6006–07. And this, the district court concluded, required the government to prove that Elfenbein's interpretation of the ambiguous guidance was unreasonable. Recognizing that the government would appeal, and in case we disagreed, the district court also conditionally granted Elfenbein's motion for a new trial. *See generally United States v. Elfenbein*, 708 F. Supp. 3d 621 (D. Md. 2023).

The government now appeals both decisions.[4]

II. Discussion

In defending the judgment below, and thus attacking the jury's verdict, Elfenbein faces an uphill climb. We often reiterate that judges must tread carefully around juries.

[4] We have jurisdiction under 28 U.S.C. § 1291 and 18 U.S.C. § 3731.

For "the best method of trial, that is possible," is trial "by a jury." 1 Matthew Hale, *Pleas of the Crown* 33 (1736). One reason why is that juries afford the accused what Joseph Story called a "double security": first "against the prejudices of judges," and second "against the passions of the multitude, who may demand their victim with a clamorous precipitancy." 3 *Commentaries on the Constitution of the United States* 653 (1833).

More prosaically, we have long recognized that juries are often better at weighing evidence than judges. Juries are valuable for "[t]heir sound common sense, brought to bear upon the consideration of testimony." *Dunlop v. United States*, 165 U.S. 486, 500 (1897). We rely on "the commonsense judgment of a group of laymen" not just because of "the community participation and shared responsibility that results," *Williams v. Florida*, 399 U.S. 78, 100 (1970), but because that group's "practical knowledge of men and the ways of men" helps find the truth, *United States v. Scheffer*, 523 U.S. 303, 313 (1998) (quoting *Aetna Life Ins. v. Ward*, 140 U.S. 76, 88 (1891)). Even in a case like this one that appears to present technical questions, there is little substitute for the jury's practical wisdom.

For these reasons, we seldom interfere with a jury's verdict. When a jury acquits, that decision is final. And when it convicts, we "require[e] only that jurors 'dr[e]w reasonable inferences from basic facts to ultimate facts.'" *Coleman v. Johnson*, 566 U.S. 650, 655 (2012) (quoting *Jackson v. Virginia*, 443 U.S. 307, 319 (1979)). This rule preserves the jury's preeminent role in criminal justice and keeps appellate courts out of the business of "fine-grained factual parsing." *Id.*

Though this deference is not limitless, it does cover Elfenbein's case. Trial courts rightly "impress[] upon the factfinder the need to reach a subjective state of near certitude"

to convict. *Jackson*, 443 U.S. at 315. And after trial, appellate courts only confirm that conviction was *possible* based on the evidence—no matter how we would have decided the case if we were in the jury's shoes. *Id.* at 318–19. Probing the verdict "only to the extent necessary," we ask "whether, after viewing the evidence in the light most favorable to the prosecution, *any* rational trier of fact could have" convicted. *Id.* at 319. Although *Jackson* set this rule in the context of postconviction review, it also applies in cases like Elfenbein's, where we review a judgment of acquittal. *See* Fed. R. Crim. P. 29 (authorizing district courts to grant judgments of acquittal); *United States v. Rafiekian*, 991 F.3d 529, 544 (4th Cir. 2021) ("We review that ruling de novo."). Because we think the government met that low bar here, we reverse the district court's Rule 29 decision.

Even so, when they are not convinced that the evidence was one-sided, district courts have some discretion to order a new trial "if the interest of justice so requires." Fed. R. Crim. P. 33(a). But they must use this power "sparingly" and "only when the evidence weighs so heavily against the verdict that it would be unjust to enter judgment." *United States v. Millender*, 970 F.3d 523, 531 (4th Cir. 2020) (cleaned up) (quoting *United States v. Arrington*, 757 F.2d 1484, 1485 (4th Cir. 1985)). Since this decision is committed to the district court's discretion, we review it only to be sure that discretion was not abused. *Id.*; *see also United States v. Fulton*, 136 F.4th 185, 191 (4th Cir. 2025) (observing that district courts have "wider latitude" in handling Rule 33 motions than Rule 29 motions). We see the evidence differently than the district court, but we detect no abuse of discretion. So we affirm the district court's Rule 33 decision.

A. A Reasonable Jury Could Have Convicted Elfenbein Of Fraud

The federal-healthcare-fraud statute makes it a crime to "knowingly and willfully execute[], or attempt[] to execute, a scheme or artifice . . . to defraud any health care benefit program." 18 U.S.C. § 1347(a), -(1). There are many ways to commit this crime. The statute forbids people to "obtain, by means of false or fraudulent pretenses, representations, or promises, any of the money or property owned by, or under the custody or control of, any health care benefit program, in connection with the delivery of or payment for health care benefits, items, or services." *Id.* § 1347(a)(2).

But the statute doesn't define "defraud" or "fraudulent." "'[F]raud' connotes deception or trickery generally," yet "the term is difficult to define more precisely." *Husky Int'l Elecs., Inc. v. Ritz*, 578 U.S. 355, 360 (2016). To resolve this indeterminacy, we have held that Congress in § 1347 incorporated the "common-law understanding of fraud." *United States v. Perry*, 757 F.3d 166, 176 (4th Cir. 2014) (quoting *United States v. Colton*, 231 F.3d 890, 898 (4th Cir. 2000)); *see also Universal Health Servs. v. United States ex rel. Escobar*, 579 U.S. 176, 187 (2016) ("[T]he term 'fraudulent' is a paradigmatic example of a statutory term that incorporates the common-law meaning of fraud."). This concept "includes acts taken to conceal, create a false impression, mislead, or otherwise deceive in order to prevent the other party from acquiring material information." *Perry*, 757 F.3d at 176 (cleaned up) (quoting *Colton*, 231 F.3d at 898); *see also* Restatement (Second) of Torts § 550 (A.L.I. 1977). So if someone knowingly submits false or misleading claims for payment to a healthcare program, that conduct violates § 1347. *See United States v. McLean*, 715 F.3d 129, 137–38 (4th Cir. 2013).

The government tried to prove that Elfenbein did that in two ways. First, the upcoding theory: That Elfenbein allegedly overbilled insurers by tacking high-level codes onto simple, low-level services. Second, the false-documentation theory: That Elfenbein allegedly supported those too-high codes with fake medical reports describing services his clinic never provided. We think the jury had enough evidence to accept both theories.[5]

1. A jury could reasonably have concluded that Elfenbein's clinic submitted false or misleading reports

Start with the upcoding theory. As explained, what code accurately describes a patient visit depends on the meanings assigned to each code by the *CPT Manual*. But the *Manual*, like any reference book, is both descriptive and prescriptive. That is, it reflects the ways that medical professionals behave while also guiding that behavior. So—as we will explain in more detail soon—we interpret its terms not just by reference to other terms in the *Manual* but also by the evidence, presented at Elfenbein's trial, of how expert coders and medical practitioners used the codes.

But first, some background. Until the pandemic, a level-four E/M code for a new patient required three elements: a comprehensive medical history, a comprehensive examination, and moderately complex medical decisionmaking. Am. Med. Ass'n, *2020 CPT Manual* 13. For established patients, though, the provider needed just two of those three elements, and the history and decisionmaking only needed to be "detailed" rather

[5] As we read it, the district court's decision did not rest on whether the government proved *scienter* or had enough evidence that Elfenbein's conduct amounted to a "scheme or artifice" under § 1347. Neither do the parties' submissions on appeal. So we do not address these questions. We instead focus on the question that was dispositive below: whether the government had enough evidence that Elfenbein's statements were false.

than "comprehensive." *Id.* at 14. Normally, as Quindoza testified at trial, this meant a level four E/M code would not be appropriate unless the provider spent a meaningful amount of time with the patient. *See id.*

The pandemic changed that. In early 2020, CMS published an interim final rule that "remove[d] any requirements regarding documentation of history and/or physical exam in the medical record." Policy and Regulatory Revisions in Response to the COVID-19 Public Health Emergency, 85 Fed. Reg. 19230, 19269 (Apr. 6, 2020). Under that rule, a provider could select the appropriate E/M code based only on how much medical decisionmaking a visit involved. *See id.* Following CMS's lead, the 2021 edition of the *CPT Manual* adopted the same rule. Under the *2021 CPT Manual*, "the extent of history and physical examination is not an element in selection of the level of . . . code[]." Am. Med. Ass'n, *2021 CPT Manual* 12. Providers still needed to perform "medically appropriate history and/or examination." *Id.* at 19. But as Elfenbein's witness Miscoe explained, that requirement was "not part of the [code] scoring elements." J.A. 1782.

Under this pandemic-era system, medical decisionmaking determined what level an E/M visit should get. In general, "medical decisionmaking" is what it sounds like. It depends on (1) "[t]he number and complexity of problem(s) that are addressed during the encounter," (2) "[t]he amount and/or complexity of the data to be reviewed and analyzed," and (3) "[t]he risk of complications and/or morbidity or mortality." *2021 CPT Manual* 14. To make this list more concrete, the *Manual* includes a table. Consider the relevant part:

Code	Level of MDM (Based on 2 out of 3 Elements of MDM)	Number and Complexity of Problems Addressed	Elements of Medical Decision Making Amount and/or Complexity of Data to be Reviewed and Analyzed **Each unique test, order, or document contributes to the combination of 2 or combination of 3 in Category 1 below.*
99211	N/A	N/A	N/A
99202 99212	Straightforward	**Minimal** • **1** self-limited or minor problem	**Minimal or none**

99203
99213
	Low	**Low**
		• **2** or more self-limited or minor problems;
		or
		• **1** stable chronic illness;
		or
		• **1** acute, uncomplicated illness or injury
			Limited
			(Must meet the requirements of at least 1 of the 2 categories)
			Category 1: Tests and documents
			• **Any combination of 2 from the following:**
			  • Review of prior external note(s) from each unique source*;
			  • review of the result(s) of each unique test*;
			  • ordering of each unique test*
			or
			Category 2: Assessment requiring an independent historian(s)
			(For the categories of independent interpretation of tests and discussion of management or test interpretation, see moderate or high)

99204
99214
	Moderate	**Moderate**
		• **1** or more chronic illnesses with exacerbation, progression, or side effects of treatment;
		or
		• **2** or more stable chronic illnesses;
		or
		• **1** undiagnosed new problem with uncertain prognosis;
		or
		• **1** acute illness with systemic symptoms;
		or
		• **1** acute complicated injury
			Moderate
			(Must meet the requirements of at least 1 out of 3 categories)
			Category 1: Tests, documents, or independent historian(s)
			• **Any combination of 3 from the following**:
			  • Review of prior external note(s) from each unique source*;
			  • Review of the result(s) of each unique test*;
			  • Ordering of each unique test*;
			  • Assessment requiring an independent historian(s)
			or
			Category 2: Independent interpretation of tests
			• Independent interpretation of a test performed by another physician/other qualified health care professional (not separately reported);
			or
			Category 3: Discussion of management or test interpretation
			• Discussion of management or test interpretation with external physician/other qualified health care professional\appropriate source (not separately reported)

J.A. 4972. These rules allow a provider to code a visit at level four if the medical decisionmaking is moderately complex in at least two respects relevant here: the problems addressed and the data reviewed. So if the provider met the criteria in Row 4, Column 4 (reviewing moderately complex data) *and* Row 4, Column 3 (addressing a moderately complex problem), then the visit should have been coded at level four.

Elfenbein argues that his coding met this standard. For each patient at issue, he says that he ordered two unique tests (rapid and PCR) and reviewed the results. And this, he says, lets him check the data-reviewed box (Column 4, Row 4, Bullet 3).[6] Here, Elfenbein's theory is hard to argue with: He offered evidence that he ordered and reviewed two tests per patient, just as level-four data analysis requires.

[6] And note the asterisk in Column 4, Row 1, which says that "each unique test . . . contributes to the combination of 2 . . . in Category 1 below."

As for the problems-addressed box, for each patient, Elfenbein says he confronted an "undiagnosed new problem with uncertain prognosis." (Column 3, Row 4, Category 1.) But the government disputed that point, and both sides offered competing evidence on how to define this phrase and whether the five patient visits charged met the definition.

Elfenbein's defense thus rested on whether he had the better of that debate—whether his patients presented undiagnosed new problems with uncertain prognoses. Helpfully, the *Manual* defines this not-so-clear phrase: It means "[a] problem in the differential diagnosis that represents a condition likely to result in a high risk of morbidity without treatment." *2021 CPT Manual* 13. Morbidity, in turn, refers to illnesses of "substantial duration during which function is limited, quality of life is impaired, or there is organ damage . . . despite treatment." *Id.* at 14. And when it comes to nontechnical terms like "high risk," the *Manual* adds that "clinicians apply common language usage meanings." *Id.*[7] So the question the jury needed to answer was whether COVID-19 was, as a matter of clinicians' usage in 2021, "a condition likely to result in a high risk of morbidity without treatment." *Id.* at 13.

[7] The district court denied this premise. It concluded that the *Manual*'s terms do not reflect common usage but are instead terms of art, and thus rejected lay testimony as not probative of their meaning. We see things differently. True enough, the *Manual* uses many terms not well known to laypersons. But the *Manual* also says that terms like "high risk" and "low risk" are to be assigned their common meanings. *2021 CPT Manual* 14. And as for terms that may indeed bear different meanings within the medical community, we struggle to see why practitioners cannot helpfully testify about their firsthand knowledge of how they and their colleagues use those words within that community—whether or not they also testify as experts.

Ample evidence let the jury conclude that it was not. To begin, consider one "example" of such a "condition" offered by the *Manual*: "a lump in the breast." *2021 CPT Manual* 13. This presents a differential diagnosis: The lump may be cancerous, or it may be benign. Put another way, the lump is a symptom with several possible causes, including breast cancer. And intuitively enough, breast cancer is "likely to result in a high risk of morbidity without treatment." *Id.* Because one leg of the differential diagnosis poses this threat, testing the lump involves moderate medical decisionmaking.

Unlike cancer, plenty of evidence suggested that for most patients, COVID-19 did not pose a high risk of morbidity without treatment. True, COVID was scary to many and dangerous to some. But Elfenbein himself testified on direct that he weighed the risks associated with COVID-19 and concluded that "for the vast majority of our COVID patients . . . it was very low[,] minimal or low risk." J.A. 2239. That aligned with what Elfenbein saw as his clinic's purpose: To handle the "simple and straightforward" task of testing patients for COVID-19, J.A. 1516, not to "solve complex medical issues," J.A. 1514. A jury could have found these descriptions inconsistent with a pitched battle against a high risk of morbidity.

And the treatments Elfenbein prescribed matched his low-risk assessment. In his words, the management plan was "minimal." J.A. 2232. For asymptomatic patients, Elfenbein recommended rest, hydration, and Tylenol. And he seems to have prescribed the same treatments to all the patients whose visits the government charged, even though some of those patients *did* present symptoms. Armed with a dose of common sense, the

jury could reasonably have concluded that these treatments do not suggest a high morbidity risk. *See* J.A. 363 (instructing that jury to use its "reason, judgment, and common sense").[8]

Elfenbein's expert confirmed that these treatments did not correspond to high-risk diseases. During his testimony, Miscoe described the requirements for a level-four code, and also the requirements for other codes. He explained that a level-two code is appropriate when the provider deals with "a self-limited or minor problem." J.A. 1869–70. Miscoe also said that "the key" to determining whether a problem is self-limited or minor "is what's the management." J.A. 1901. And he agreed that treatments like rest and over-the-counter drugs matched level two, not level four. By comparing that testimony with the records of Elfenbein's treatments, the jury could have concluded that COVID-19 was not—at least for most patients—a high-risk condition.[9]

[8] To be sure, some of Elfenbein's testimony cut the other way. For instance, he argued that COVID-19 carried an uncertain prognosis because "[p]eople were dying" and "[t]he world was shut down." J.A. 2230–31. And he added that for symptomatic patients, "by definition," those patients faced "a threat to life or bodily function." J.A. 2240. Maybe, but maybe not. As Elfenbein's coding specialist testified, "unless the patient's in respiratory distress I don't know I can make a case for threat to life or limb." J.A. 879. And the visits the government charged were for patients who reported few if any symptoms. Either way, with conflicting testimony, the jury was free to choose what to credit.

[9] Miscoe also opined that in 2021, "there was absolutely no certainty" that COVID-19 would affect a patient, or that the resulting sickness would "resolve with appropriate treatment." J.A. 1901. This may have been sufficient evidence for the jury to conclude that COVID-19 *was* a condition with uncertain prognosis. But it does not contradict Elfenbein's testimony about the treatments he actually provided—or negate Miscoe's opinion that those treatments matched level two, not level four. And in all events, the jury did not have to credit Miscoe's no-certainty testimony.

Staff in Elfenbein's clinic agreed. In their "common language usage," *2021 CPT Manual* 14, COVID-19 tests didn't count as level-four visits. The jury heard testimony along these lines from Elfenbein's coding specialist, Cathy Raymond, and two care providers. True, not every witness explained why he or she thought a level-four code was too high. But the employees' shared concern about using level-four codes was still evidence of common usage that could have played a role in how the jury interpreted the codes.

Last, the jury heard from an auditor who checked Elfenbein's books on behalf of a private insurer, CareFirst. The audit detected multiple problems, including "[i]mproper coding" for COVID-19 tests. J.A. 1176. As CareFirst saw things, COVID-19 testing visits were "basic, low-level evaluation and management"—warranting level two, or perhaps level three, but not level four. J.A. 1199. So for the clinic's level-four codes, the audit "yielded an error rate of 100 percent." J.A. 1200. To be sure, the audit was imperfect. CareFirst could not access all of the clinic's records, and perhaps the audit's result would have changed had the auditors seen everything. But this does not undercut the auditor's testimony about how CareFirst understood the codes. The mere fact that CareFirst saw COVID-19 tests as simple, low-level care supported the jury's conclusion that doing the tests did not involve moderately complex decisionmaking.

The jury did not hear evidence that anything extra added complexity to the five charged visits. Each patient was tested for COVID-19, some because they had symptoms and others just because they thought they might have been exposed or needed a negative test result to go about their lives. None of the patients reported exceptional symptoms.

Nobody testified that any of the patients had risk factors (like age or other health problems) that suggested that they faced a substantial risk from COVID-19. Each patient was prescribed either no treatment or basic treatment—rest, hydration, and over-the-counter medication. Still, Elfenbein's clinic coded all the visits at level four. Based on the rest of the evidence it had seen, the jury was entitled to conclude that it was fraudulent to code these visits at level four.

Next, consider the government's false-documentation theory. Recall, one way to commit healthcare fraud is to submit false records in support of a claim for payment. *See, e.g.*, *McLean*, 715 F.3d at 137–38. Like billing codes, the medical records supporting a bill are sent to insurers as part of a request for payment. And whether the insurer pays depends on both whether the code accurately describes the services *and* whether those services were actually rendered and "medically necessary." J.A. 365–66. So as with the codes, the jury could have convicted Elfenbein if it decided that the medical records were materially false or misleading.[10]

[10] Of course, a trivial mistake would not usually support a fraud conviction. "[T]he common law has long embraced . . . materiality . . . as the principled basis for distinguishing everyday misstatements from actionable fraud." *Kousisis v. United States*, 145 S. Ct. 1382, 1396 (2025). This means the false statement must have "a natural tendency to influence, or [be] capable of influencing, the decision of the decisionmaking body to which it is addressed." *Neder v. United States*, 527 U.S. 1, 16 (1999) (quoting *United States v. Gaudin*, 515 U.S. 506, 509 (1995)). And equally important, we require *mens rea*. *McLean*, 715 F.3d at 137. An innocent or immaterial error would not meet this standard.

In accord with these principles, the jury was instructed that it could only convict if the alleged "fraudulent representation" was "material," or "one that would reasonably be expected to be of concern to a reasonable and prudent person in relying upon the representation or statement in making a decision." J.A. 2456. And Elfenbein does not (Continued)

The jury had enough evidence to reach that conclusion. One patient, A.H., testified that when she went to the clinic, she didn't receive any treatment, testing, or examination besides a nostril swab and a short oral questionnaire. But the medical record Elfenbein's clinic sent to her insurer said that they had checked her temperature, pulse, oxygen saturation, and respiration, even though A.H. testified that no one had actually done so. So too with another patient, J.J., whose records showed that Elfenbein's clinic checked her vitals, even though she testified that it did no such thing. Another patient, S.T., testified that the clinic never called her to report her test results. But her records reflected that someone from the clinic "spoke at length over the phone [with S.T.] about [her] PCR results and what it means for [her]," and also "discussed [her] overall health and well being." J.A. 1005–06; *see also* J.A. 674–75, 3647 (indicating a similar results-call mismatch as to A.H.). Yet another patient's record said, "the pharynx is without exudates." J.A. 1347. But the provider who treated the patient testified that she "did not do all of" the checks indicated by the record, much less anything that would tell her about exudates in the pharynx. *Id.*[11]

In sum, we disagree with the district court's view of the evidence. As we read the record, the jury had enough evidence to convict Elfenbein.

---

dispute materiality—or that if the records were false in the way the government contends, he got more money from insurers than he should have and did so by billing too high. *See Ciminelli v. United States*, 598 U.S. 306, 312 (2023) (explaining that "money or property" must be "an object of the[] fraud" (quotation omitted)).

[11] Some evidence suggested that these discrepancies were caused by the clinic's use of templates that automatically populated the results of physical exams. But this cuts against Elfenbein, not for him. If the problems resulted from clinic-wide policies rather than individual corner-cutting nurses, then it seems more likely that Elfenbein knew about—or created—the discrepancies.

2. Elfenbein's replies are unpersuasive

Elfenbein protests that despite all this, his statements weren't false. This, he says, is because concepts like "undiagnosed new problem of uncertain prognosis" are open to interpretation. On Elfenbein's view, the CPT codes are so open-textured that any particular use of the codes could seldom, if ever, be meaningfully false. One care provider might in her judgment deny that COVID-19 counted as an undiagnosed new problem with uncertain prognosis, but others might disagree. Therefore, says Elfenbein, neither provider can be objectively wrong.

In the abstract, it's true that reasonable people could disagree about what the codes mean. (For evidence of that, just consider the dueling experts below.) But the possibility of reasonable disagreement doesn't rule out falsity. In the main, "ambiguity does not preclude" the possibility that words and phrases have a "correct meaning"—"or, at least," that people can "becom[e] aware of a substantial likelihood of the terms' correct meaning." *United States ex rel. Schutte v. SuperValu Inc.*, 598 U.S. 739, 753 (2023).

Schutte makes for a good example. There, pharmacies seeking reimbursement from Medicare and Medicaid programs were required by regulation to charge their "usual and customary" prices to customers using private insurance plan sponsors. *Id.* at 745. Instead, they charged such customers higher-than-usual prices. For although the pharmacies charged those customers their sticker prices, in reality, "more than 80%" of noninsurance sales were made at steep discounts. *Id.* at 746. In court, the pharmacies argued that they couldn't know the claims were false because the claims couldn't *be* false—that people

could reasonably disagree about whether the "usual and customary" price was the nominal default or instead the one most often actually charged.

The Supreme Court disagreed. It began with an analogy, "a hypothetical driver who sees a road sign that says 'Drive Only Reasonable Speeds.' That driver, without any more information, might have no way of knowing what speeds are reasonable and what speeds are too fast." *Id.* at 753. But in context, a vague term like "reasonable" can acquire real meaning. If a police officer told the driver that "speeds over 50 mph are unreasonable," and the driver saw "that all the other cars around him are going only 48 mph," then "the driver might know that 'Reasonable Speeds' are anything under 50 mph." *Id.* Applying this principle to the pharmacies, the Court pointed out that insurers told the pharmacies "the phrase 'usual and customary' referred to their discounted prices." *Id.* at 754. That evidence of common usage was enough, at least in principle, for the pharmacies to "actually kn[o]w what the phrase meant." *Id.*[12] In other words, language that doesn't mean much on its own can acquire meaning from context. And once it does, people can use that language to tell the truth—or not.

Reflecting this idea that vague words and phrases can acquire truth values from context, we have held that only "fundamentally ambiguous" language cannot "form the

[12] To be sure, *Schutte* focused on *scienter*, and thus *knowledge* of falsity, not falsity itself. But of course, the Court's opinion could have been much shorter if it thought the claims at issue were too ambiguous to admit of truth values. If a statement has no truth value, then someone cannot *know* it is false. *Cf.* Edmund L. Gettier, *Is Justified True Belief Knowledge?*, Analysis, June 1963, at 121. *Schutte* thus takes as a premise that when customary usage determines a phrase's meaning, that phrase's "facial ambiguity . . . does not by itself preclude" the falsity of statements using it. 598 U.S. at 754.

basis for a false statement." *United States v. Sarwari*, 669 F.3d 401, 407 (4th Cir. 2012). This admittedly slippery category—"the exception, not the rule"—covers language without "a meaning about which men of ordinary intellect could agree." *Id.* (first quoting *United States v. Farmer*, 137 F.3d 1265, 1269 (10th Cir. 1998); then quoting *United States v. Lighte*, 782 F.2d 367, 375 (2d Cir. 1986)). It is not enough that "the words used . . . have different meanings in different situations." *Id.* (quoting *Lighte*, 782 F.2d at 375). To fall into this category, the words must be intractable. So long as it is "reasonable to expect a defendant to have understood the terms used" in their context, they can support a fraud conviction. *Id.* (quoting *United States v. Long*, 534 F.2d 1097, 1101 (3d Cir. 1976)).[13]

This is not a case of fundamental ambiguity. The parties' dispute about the various meanings one might attach to the phrase "undiagnosed new problem with uncertain prognosis" is more "semantic" than metaphysical. *Id.* at 408. Nurses, coding specialists, two expert witnesses, the *CPT Manual*, and various federal regulations all assert that phrases like these can be assigned more or less definite meaning. That does not mean they contain no vagueness. But the American health insurance system depends on the proposition that it is possible to categorize degrees of medical decisionmaking. And to serve that end, the *Manual* tells us how to deal with linguistic indeterminacy: Follow the

[13] When *Sarwari* speaks of *a* defendant, it does not mean *the* defendant. Whether a statement is false and whether the person on trial *knows* it to be false are different questions. And whether a word or phrase is fundamentally ambiguous goes only to falsity—not knowledge. The reason we ask whether *a* defendant would reasonably understand the words is that we need to know whether the words mean something to reasonable people. If they do, then it is possible to use those words to make a false statement. If they do not—in other words, if they are fundamentally ambiguous—then it is *not* possible to use those words to make a false statement.

usage that prevails among care providers on the ground. The *Manual* thus presumes—and we agree—that "an average person would understand" that it is both possible and forbidden to make a false statement by misapplying the codes. *McLean*, 715 F.3d at 137. Given all this, we will not adopt the skepticism Elfenbein urges.

Retreating to less metaphysical ground, Elfenbein counters that even if the *possibility* of reasonable disagreement doesn't get him off the hook, *his* interpretations were reasonable—and that was enough. In other words, if the *CPT Manual* is not fundamentally ambiguous, Elfenbein urges that it is still somewhat vague. And this, he says, creates a safe harbor. On Elfenbein's proposed rule, a statement that lines up with any reasonable interpretation of the terms it uses is not false.

Once again, we do not deny the minor premise. Reasonable people could indeed interpret the *CPT Manual* differently. But this is what juries are for. In criminal proceedings, "if the evidence supports different, reasonable interpretations" of the relevant facts, then "the jury decides which interpretation to believe." *United States v. Burgos,* 94 F.3d 849, 862 (4th Cir. 1996) (en banc) (cleaned up) (quoting *United States v. Murphy*, 35 F.3d 143, 148 (4th Cir. 1994)). And in fraud cases where liability depends on the falsity of words that "admit[] of two reasonable interpretations," it is the jury's job to decide—based on the evidence—which interpretation is better. *Sarwari*, 669 F.3d at 409 (quoting

Farmer, 137 F.3d at 1269). The jury did that here and evidently found Elfenbein's interpretation lacking.[14]

To this, Elfenbein replies that the jury had too little evidence to decide what a level-four code meant or whether COVID-19 testing met the definition. In his view, the jury

---

[14] The district court's contrary conclusion relied on an out-of-circuit case, *United States v. Harra*, 985 F.3d 196 (3d Cir. 2021). The Third Circuit there decided that a statement is false, and so supports fraud liability, only if it contradicts *all* reasonable interpretations of the words it uses—not just the best interpretation. *See* 985 F.3d at 204. We too used to take this view. *See United States v. Race*, 632 F.2d 1114, 1120 (4th Cir. 1980) ("[O]ne cannot be found guilty of a false statement under a contract . . . when his statement is within a reasonable construction of the contract."). But we have since "disavow[ed] the *Race* dicta." *Sarwari*, 669 F.3d at 407 n.3.

The district court read *Sarwari* narrowly, to disavow that rule only in cases reducible to yes-or-no questions that the defendant claims he answered with literal truth. *See Elfenbein*, 708 F. Supp. 3d at 661 n.20. We think this reading too stingy. Although *Sarwari* was "a case about" literal truth, "this does not mean it was *only* a case about" literal truth. *City of Martinsville v. Express Scripts, Inc.*, 128 F.4th 265, 270–71 (4th Cir. 2025). *Sarwari* reasoned that literal truth is "a defense *only* if a defendant's statement is literally true, not if simply a 'reasonable construction.'" 669 F.3d at 407 n.3. In so doing, *Sarwari* made clear that a false-statement conviction is possible even when words are "susceptible to multiple interpretations." *Id.* at 407.

In other words, whether a statement is true depends on whether it tracks the best interpretation of the words it uses, not just a reasonable interpretation. It is only when there *is* no best interpretation, because the words are fundamentally ambiguous, that the defendant's statements cannot be either true or false. This leaves us with two types of cases: (1) cases where the relevant language is fundamentally ambiguous and false-statement conviction is therefore impossible, and (2) cases where the language is somewhat ambiguous but not fundamentally so, which means the defendant can be convicted if his statements contradicted the best interpretation of the language at issue.

To be sure, under *Sarwari*, what counts as the best interpretation is up to the factfinder. *See* 669 F.3d at 407 (explaining that "the fact finder determines" how to interpret the language at issue and that "[a]n appellate court's only role . . . is to assess the sufficiency of the evidence"). We thus agree with our good colleagues up north that "the construction of an arguably ambiguous question or reporting requirement" is a matter for the jury. *Harra*, 985 F.3d at 216 (quotation omitted). But our precedent forbids us to join them in thinking that fraud defendants benefit from *Chevron*-like safe harbors. *See id.* at 218 (quoting *Chevron v. Nat. Res. Def. Council*, 467 U.S. 837, 843 n.9 (1984)).

could only decide that COVID-19 tests fell short of level four if an expert said so expressly. Once again, we disagree. Experts may opine on ultimate issues, but that does not mean they must. Until fifty years ago, experts *could not* opine on "issues that the jury must resolve to decide the case." *Diaz v. United States*, 602 U.S. 526, 531 (2024). The rationale for this rule was that letting an expert state a view on the very question the jury must decide—whether an injury was caused by malpractice, who fired the killing shot, or whether a statement was false—would leave the jury "with no other duty but that of recording the finding of the witness." *Id.* at 532 (quoting *Chicago & Alton Ry. Co. v. Springfield & N. W. R.R. Co.*, 67 Ill. 142, 145 (1873)). Though we have since abolished this rule, courts often admit—and sometimes prefer—expert testimony that offers background information yet does not close the loop. *See United States v. Campbell*, 963 F.3d 309, 313–14 (4th Cir. 2020); *United States v. Offill*, 666 F.3d 168, 174 (4th Cir. 2011).

For good reason. The purpose of experts in federal trials is not to decide technical questions; it is to bring the unfamiliar aspects of those questions within the jury's ken so that the jury can decide them. As the advisory committee explained when it abolished the so-called ultimate-issue rule, once an expert has explained the technical or conceptual framework the jury must apply, *applying it* is normally the jury's job. *See* Fed. R. Evid. 702 advisory committee's note to 1972 amendment ("[A]n expert on the stand may give a dissertation or exposition of scientific or other principles relevant to the case, leaving the trier of fact to apply them to the facts."). So long as experts have said enough to bring

technical questions within the jury's competence, the jury may answer them—whether or not an expert also says what answer he would give.[15]

The jury had enough information about CPT codes to apply them here. Although no expert said in so many words that level-four codes were inappropriate for COVID-19 tests, the experts *did* explain what the codes meant in general. For instance, Miscoe testified that under the *CPT Manual*, level-four codes go with high-risk problems. By comparing this with Elfenbein's testimony that COVID-19 generally posed a *low* risk, the jury could conclude that the codes were too high.

Along similar lines, Miscoe testified that the treatment prescribed by the physician was "the key" to determining whether a problem warranted a level-four code. J.A. 1901. And as explained, Miscoe opined that basic treatment like rest and over-the-counter pain medication did not fit the bill. With this background in hand, the jury faced a straightforward task: Cross-reference Miscoe's explanation about what treatment indicated level-four conditions with evidence about the treatments Elfenbein prescribed. The jury was free to, and seemingly did, credit Miscoe's testimony about the *meaning* of

[15] To be sure, some applications may be sufficiently complex that an expert must do more than explain a general principle. If an expert testifies about what a CT scan is generally, we doubt that a jury could then examine a scan and decide whether a mass looks cancerous. But this example does not change the general principle that an expert need only give the jury enough information so that it has the capacity to decide technical questions. It only illustrates a corollary: The more detail an expert gives the jury about the relevant conceptual framework, the less likely it will be that the expert must also opine on how the framework applies to the facts in question because the jury will be able to perform this step itself. Conversely, if an expert provides less detail about the framework, application testimony may be more important.

the codes while disagreeing with Miscoe about whether Elfenbein's *use* of the codes fit that meaning. We see no error there.[16]

Elfenbein last invokes the jury standard itself. Whatever we make of these problems, he insists, they at least created reasonable doubt below. And given that, he asserts, the jury could not convict. But this reply confuses the rule applied by factfinders with the rule applied by judges reviewing those findings.

The venerable words "beyond a reasonable doubt" describe how juries must reach their conclusions. They mean that jurors may only vote to convict if they are sure the defendant is guilty. This rule is important: It is "bedrock" that "lies at the foundation of . . . our criminal law" and has "constitutional stature." *In re Winship*, 397 U.S. 358, 363–64 (1970) (quoting *Coffin v. United States*, 156 U.S. 432, 453 (1895)). But the rule is also aspirational. Absolute proof lies beyond mortal humans; "the beyond a reasonable doubt standard is itself probabilistic." *Victor v. Nebraska*, 511 U.S. 1, 14 (1994). Phrases like "reasonable doubt" are "quantitatively imprecise" because "no one has yet invented . . . a mode of measurement for the intensity of human belief." *Winship*, 397 U.S. at 369 (Harlan, J., concurring) (quoting 9 John Wigmore, *Evidence* 325 (3d ed. 1940)).

[16] In so concluding, we have assumed without deciding that the narrow question on which falsity rested—whether COVID-19 counted in 2021 as a problem "with uncertain prognosis"—*was* a question about which the jury needed expert testimony. But this premise is not obvious. Expert testimony is required only if some matter in dispute lies outside lay common knowledge. It would not, for instance, be necessary in order for a jury to conclude that the flu has an uncertain prognosis—that it sometimes but seldom kills and usually causes only moderate symptoms. By contrast, a jury may well need an expert's help to make the same call about a rarer condition like cystic fibrosis. And of course, the scope of common knowledge changes over time. In 2023, when Elfenbein was tried, the dangers of COVID-19 may have fallen into this category.

Indeed, the aspiration is part of the point. By requiring "a subjective state of near certitude of the guilt of the accused, the standard symbolizes the significance that our society attaches to the criminal sanction and thus to liberty itself." *Jackson*, 443 U.S. at 315.

As important as these lofty words are, their subjective focus should make clear that they do not describe the way district courts evaluate Rule 29 motions—or the way we review jury verdicts on appeal. In those contexts, we have explained, deference to juries requires nearly the opposite rule: While a jury must not convict if it could reasonably acquit, a judge must not order the jury to acquit if it could reasonably convict. *See Jackson*, 443 U.S. at 318; *accord United States v. Rafiekian* (*Rafiekian II*), 68 F.4th 177, 186 (4th Cir. 2023). Sometimes, the difference between conviction and acquittal comes to whether the jury believes a single witness. *See Carmell v. Texas*, 529 U.S. 513, 541–42 & n.30 (2000); John H. Wigmore, *Required Numbers of Witnesses; A Brief History of the Numerical System in England*, 15 Harv. L. Rev. 83, 93 (1901). It is not for us to second-guess the jury's belief. *Burgos*, 94 F.3d at 860–61. Or, closer to this case, the jury's decision may rest on how it interprets cryptic testimony. That, too, is the jury's job—not ours. *See Sarwari*, 669 F.3d at 409. On a cold record, situations like these may tempt us to wonder whether the jury truly was doubt-free. But that is not the question we're supposed to ask. So long as anyone could look at the evidence and reasonably conclude that the defendant committed the crime, we leave the defendant's fate with the jury. In

more concrete terms, so long as a witness testified to the existence of each element, and the jury was not obligated to discredit that witness, we will not disturb the jury's verdict.[17]

B. Even So, Granting A New Trial Was Not An Abuse Of Discretion

As explained, the jury had enough evidence to convict Elfenbein. But it got that evidence in an unusual way. At the close of the government's case-in-chief, the jury had little of the key evidence—no clear, general explanation about what level-four codes required or the *CPT Manual*'s terms meant; no testimony about how Elfenbein and his staff used terms like "low risk"; and only partial information about the treatments Elfenbein prescribed and his reasons for doing so. Most of that information came from Miscoe (Elfenbein's expert) and Elfenbein himself, who took the stand in his defense.

That fact makes no difference to whether the jury could convict. And on first review, we might agree with the jury's weighing of the evidence. But the district court is owed deference in granting a new trial under Rule 33. *See Rafiekian II*, 68 F.4th at 186–87. By that deferential standard, the weaknesses in the government's case-in-chief lead us to find no abuse of discretion in the district court's decision to try the case again.

[17] We do not take Elfenbein to separately argue that he lacked *scienter* even if the codes were false. But we note that we see the evidence as sufficient here too. In fraud cases, knowledge must often be inferred "from the totality of the circumstances," not "proven by direct evidence." *McLean*, 715 F.3d at 138 (quoting *United States v. Harvey*, 532 F.3d 326, 334 (4th Cir. 2008)). The jury could have made that inference here. Elfenbein said many times that COVID-19 test visits were neither complex nor lengthy. And he was told by staff in his clinic and independent auditors that his codes were too high. If the codes were indeed improper—which again depends on the common usage of medical care providers—then the jury could infer that Elfenbein knew it.

* * *

Sometimes, complex cases reduce to simple questions. Overall, the government's evidence against Elfenbein was thin. But Elfenbein's expert testified about the high risk and significant treatment that warranted level-four codes, and Elfenbein testified that he neither saw the risk nor prescribed the treatments. Given this, the jury could have convicted. All the same, the district court was within bounds to order a do-over.

AFFIRMED IN PART, REVERSED IN PART, AND REMANDED