

ELEMENTS

HOW A SHARP-EYED SCIENTIST BECAME BIOLOGY'S IMAGE DETECTIVE

Using just her eyes and memory, Elisabeth Bik has single-handedly identified thousands of studies containing potentially doctored scientific images.

By Ingfei Chen

June 23, 2021



Illustration by Kailey Whitman



Save this story

In June of 2013, Elisabeth Bik, a microbiologist, grew curious about the subject of plagiarism. She had read that scientific dishonesty was a growing problem, and she idly wondered if her work might have been stolen by others. One day, she pasted a sentence from one of her scientific papers into the Google Scholar search engine. She found that several of her sentences had been copied, without permission, in an obscure online book. She pasted a few more sentences from the same book chapter into the search box, and discovered that some of them had been purloined from other scientists' writings.

Bik has a methodical, thorough disposition, and she analyzed the chapter over the weekend. She found that it contained text plagiarized from eighteen uncredited sources, which she categorized using color-coded highlighting. Searching out plagiarism became a kind of hobby for Bik; she began trawling Google Scholar for more cases in her off-hours, when she wasn't working as a researcher at Stanford. She soon identified thirty faked biomedical papers, some in well-respected journals. She e-mailed the publications' editors, and, within a few months, some of the articles were retracted.

In January, 2014, Bik was scrolling through a suspicious dissertation when she began glancing at the images, too. They included photographs known as Western blots, in which proteins appear as dark bands. Bik thought that she'd seen one particular protein band before—it had a fat little black dot at one end. Elsewhere in the dissertation, she found the same band flipped around and presented as if it were data from a different experiment. She kept looking, and spotted a dozen more Western blots that looked copied or subtly doctored. She learned that the thesis, written by a graduate student at Case Western Reserve University, had been published as two journal articles in 2010.

Get the Science & Technology newsletter

In-depth reporting and commentary on A.I., medicine, psychology, the climate, and more.

[Sign up](#)

By signing up, you agree to our [user agreement](#) (including [class action waiver and arbitration provisions](#)), and acknowledge our [privacy policy](#).

The presence of a flawed image in a scientific study doesn't necessarily invalidate its central observations. But it can be a sign that something is amiss. In science, images are profoundly important: every picture and graph in a scientific paper is meant to represent data supporting the authors' findings. Photographic images, in particular, aren't illustrations but the evidence itself. It seemed to Bik that duplicated or doctored images could be more damaging to science than plagiarism.

Bik decided to scan through some newly published studies in *PLOS One*, an "open access" journal in which articles are made available to the public free of charge. (The journal's nonprofit publisher charges authors article-processing fees.) She opened fifteen articles, each in its own browser tab, and began eyeballing the images without reading the text. In a few hours, she'd looked at around a hundred studies and spotted a few duplicate images. "It very quickly became addictive," Bik told me, in a marked Dutch accent. Night after night, she collected problematic articles, some with duplicate Western blots, others with copied images of cells or tissues. All had passed through peer review before being accepted. A few duplications could have been innocent—perhaps a mixup by a scientist with a folder full of files. But other images had been cloned, stretched, zoomed, rotated, or reversed. The forms and patterns in biology are endlessly unique; Bik knew that these duplications couldn't have happened by accident. Yet she didn't want to mistakenly implicate a fellow-scientist in wrongdoing. She sent polite e-mails to the journals that had published the two Case Western studies. Editors eventually

replied, promising to look into her concerns. Then six months passed with no further word. Bik was stymied.

In 2012, three scientists had created a Web site called [PubPeer](#), where researchers could discuss one another's published work. Critics objected to the fact that the site allowed anonymous comments. Still, PubPeer was moderated to prohibit unsubstantiated accusations, and, in several cases, unnamed whistle-blowers had used it to bring attention to image manipulations or statistical errors, spurring major corrections and retractions. It seemed to Bik that posting her findings online involved crossing a boundary: the traditional way to raise questions about a paper's integrity was private communication with the authors, journals, or universities. She made an anonymous account anyway. "I have concerns about some figures in this paper," she wrote, for each Case Western study. She uploaded screenshots of the image duplications, with the key areas clearly delineated by blue or red boxes, and clicked the button to submit.

Scientific publishing is a multibillion-dollar industry. In biomedicine alone, more than 1.3 million papers are published each year; in all of science, there are more than twelve thousand reputable journals. Thousands of other Web-based journals publish even the flimsiest manuscripts after sham peer review, in exchange for processing fees. In China, researchers under pressure to meet unrealistic publication quotas purchase ghostwritten papers on a black market. Meanwhile, as the Web has made it easy for journals to proliferate, professional advancement in science has increasingly depended on publishing as many studies as possible.

Around a decade ago, scientists began reckoning with the effects of this supercharged publish-or-perish system. A few cases of outright fraud—including the British study that falsely linked vaccines to autism—troubled specific scientific disciplines; in psychology, cancer research, and other fields, it was recognized that a meaningful proportion of studies had made overreaching claims and couldn't be

replicated. Reforms were introduced. Watchdog Web sites such as PubPeer and Retraction Watch sprang up, and a number of independent research-integrity detectives began unearthing cases of misconduct and sharing them through blogs, PubPeer, and on Twitter.

In March of 2019, when she was fifty-three, Bik decided to leave her job to do this detective work full time, launching a blog called Science Integrity Digest. Over the past six and a half years—while earning a bit of income from consulting and speaking, and receiving some crowdfunding—she has identified more than forty-nine hundred articles containing suspect image duplications, documenting them in a master spreadsheet. On Twitter, more than a hundred thousand people now follow her exposés.

Bik grew up with two siblings in Gouda, in the Netherlands, where her mother and physician father ran a medical practice out of their red-brick house, on a tree-lined canal. At the age of eight, Bik wanted to become an ornithologist, and spent hours with binoculars, scanning the garden for birds and recording all the species she sighted. She discovered science, earned a Ph.D. in microbiology, and moved to the United States just after 9/11, when her husband, Gerard, an optical engineer, got a job in Silicon Valley. She spent fifteen years studying the microbiome in a Stanford laboratory before moving on to the biotech industry.

When Bik first stumbled upon the image-duplication issue, a few journal editors had been writing about it, but no one had ascertained the scale of the problem. She e-mailed two prominent microbiologists, Ferric Fang and Arturo Casadevall, who had studied retractions in science publishing, introducing herself along with image duplications she'd found in *Infection and Immunity* and *mBio*—journals for which Fang and Casadevall were the editors-in-chief, respectively. The three agreed to a systematic study. Bik would screen papers in forty different journals, and Fang and Casadevall would review her findings.

In 2016, the team published their results in *mBio*. When journal editors examine questionable images, they typically use Photoshop tools that magnify, invert, stretch, or overlay pictures, but Bik does the same work mostly with her eyes and memory alone. Working at a speed of a few minutes per article, she had screened a jaw-dropping 20,621 studies. The team concluded that she was right ninety per cent of the time; the remaining ten per cent of images included some that were too low-resolution to allow for a clear determination. They reported “inappropriate” image duplications in seven hundred and eighty-two, or four per cent, of the papers; around a third of the flagged images involved simple copies, which could have been inadvertent errors, but at least half of the cases were sophisticated duplications which had likely been doctored. “Sometimes it seems almost like magic that the brain can do this,” Fang told me, of Bik’s abilities.

The trio estimated that, of the millions of published biomedical studies, tens of thousands ought to be retracted for unreliable or faked images. But adjusting the scientific record can be maddeningly slow, especially when research is lower-profile. In total, it took journal editors more than thirty months to retract the two Case Western papers that Bik had reported. In addition to contacting editors, Bik sometimes reaches out to research institutions, or to the Office of Research Integrity (O.R.I.), a government agency responsible for investigating misconduct in federally funded science. But the O.R.I. and institutions have protocols—they must obtain lab notebooks, conduct interviews, and so on—which take time to unfold.

By 2016, Bik had reported all seven hundred and eighty-two papers in the *mBio* study to journal editors (including at *PLOS One*). As of this June, two hundred and twenty-five had been corrected, twelve had been tagged with “expressions of concern,” and eighty-nine had been retracted. (Among them were five discredited studies by a cancer researcher at Pfizer, who was fired.) As far as Bik knows, fifty-eight per cent of the studies remain at large. In the past five years, she has reported problematic images in another 4,132 studies; only around fifteen per cent have

been addressed so far. (Three hundred and eighty-two have been retracted.) In only five or ten cases has she been told that authors proved her image concerns to be unfounded, she said.

Frustrated by these long timetables, Bik has transitioned to sharing more of her findings online, where journal readers can encounter them. On PubPeer, where she is the most prolific poster who uses her real name, her comments are circumspect —she writes that images are “remarkably similar” or “more similar than expected.” On Twitter, she is more performative, and often plays to a live audience.

“#ImageForensics Middle of the Night edition. Level: easy to advanced,” Bik tweeted, at 2:41 A.M. one night. She posted an array of colorful photographs that resembled abstract paintings, including a striated vista of pink and white brushstrokes (a slice of heart tissue) and a fine-grained splattering of ruby-red and white flecks (a slice of kidney). Six minutes later, a biologist in the U.K. responded: two kidney photos appeared identical, she wrote. A minute later, another user flagged the same pair, along with three lung images that looked like the same tissue sample, shifted slightly. Answers continued trickling in from others; they drew Bik-style color-coded boxes around the cloned image parts. At 3:06 A.M., Bik awarded the second user an emoji trophy for the best reply.

In Silicon Valley, Bik and her husband live in an elegant mid-century-modern ranch house with a cheerful, orange front door and a low-angled pitched roof. In the neighborhood, the residence is one of many duplicate copies sporting varying color schemes. I visited Bik just before the pandemic began. Tall, with stylish blue tortoiseshell eyeglasses and shoulder-length chestnut hair, she wore a blouse with a recurring sky-blue-and-orange floral pattern and had a penetrating, blue-eyed gaze. While Bik made tea, her husband, clad in a red fleece jacket, toasted some frozen stroopwafel cookies, from Gouda.

Playing tour guide, Bik showed off the original features of their kitchen, including its white Formica countertop, flecked with gold and black spots. “It’s random!” she

assured me—no duplications. The same could not be said of the textured gray porcelain floor tiles. When workers installed them, Bik explained, she'd asked them to rotate the pieces that were identical, so that repeats would be less noticeable. A few duplicate tiles had ended up side-by-side anyway. I couldn't see the duplication until she traced an identical wavy ridge in each tile with both of her index fingers. "Sorry—I'm, like, weird," she said, and laughed.

In her bedroom closet, Bik's shirts hung in a color gradient progressing from blacks and browns to greens and blues. Not long ago, she helped arrange her sister-in-law's enormous shoe collection by color on new storage racks; when some friends complained about the messy boxes of nuts, screws, and nails that littered their garage, Bik sorted them into little drawers. "Nothing makes me more happy," she told me. Since childhood, she has collected tortoise figurines and toys; around two thousand of them are arranged in four glass cabinets next to a blond-wood dining table. She keeps a spreadsheet tracking her turtle menagerie: there are turtles made from cowrie seashells, brass turtles, Delft blue porcelain turtles, bobble-headed turtles, turtle-shaped wood boxes with lids, and "functional" turtles (key chains, pencil sharpeners). She showed me a small stuffed animal with an eye missing: Turtle No. 1. (She has stopped adding to her collection. "I don't want it to overtake my house," she said.)

That afternoon, Bik settled at her dining table, which serves as her desk. Floor-to-ceiling windows offered a tranquil view of back-yard foliage. On her curved widescreen monitor, Bik checked her Twitter account—her bio featured a photo of a cactus garden; "That's me—prickly," she said—and then pulled up her master spreadsheet of problematic papers, which she doesn't share publicly. Each of its thousands of entries has more than twenty columns of details. She removed her glasses, set them next to a cup of chamomile tea, sat up straight, and began rapidly scanning papers from *PLOS One* with her face close to the monitor. Starting with the first study—about "leucine zipper transcription factor-like 1"—she peered at an array of Western-blot images. She took screenshots and scrutinized them in

Preview, zooming in and adjusting the contrast and brightness. (Occasionally, she uses Forensically and ImageTwin, tools that do some semi-automated photo-forensics analysis.) She moved on to a study with pink and purple cross-sections of mouse-gut tissue, then stopped on a figure with a dozen photos of translucent clumps of cells. She chuckled. “It looks like a flying rabbit,” she said, pointing at one blob.

Bik found no problems. *PLOS One* has “cleaned up their act a lot,” she said. The journal’s publisher employs a team of three editors who handle matters of publication ethics, including Bik’s cases. Renee Hoch, one of the editors, told me that the process of investigation, which entails obtaining original, raw images from the authors, and, in some cases, requesting input from external reviewers, usually takes four to six months per case. Hoch said that of the first hundred and ninety or so of Bik’s cases that the team had resolved, forty-six per cent required corrections, around forty-three per cent were retracted, and another nine per cent received “expressions of concern.” In only two of the resolved papers was nothing amiss. “In the vast majority of cases, when she raises an issue and we look into it, we agree with her assessment,” Hoch said.

Could Bik be replaced with a computer? There are arguments for the idea that automated image-scanning could be both faster and more accurate, with fewer false positives and false negatives. Hany Farid, a computer scientist and photo-forensic expert at the University of California, Berkeley, agreed that scientific misconduct is a troubling issue, but was uneasy about individual image detectives using their own judgment to publicly identify suspect images. “One wants to tread fairly lightly” when professional reputations are on the line, he told me. Farid’s reservations spring partly from a general skepticism about the accuracy of the human eye. While our visual systems excel at many tasks, such as recognizing faces, they aren’t always good at other kinds of visual discrimination. Farid sometimes provides court testimony in cases involving doctored images; his lab has designed algorithms for detecting faked photographs of everyday scenes,

and they are eighty-to-ninety-five-per-cent accurate, with false positives in roughly one in a hundred cases. Judging by courtroom standards, he is unimpressed by Bik's stats and would prefer a more rigorous assessment of her accuracy. "You can audit the algorithms," Farid said. "You can't audit her brain." He would like to see similar systems designed and validated for identifying faked or altered scientific images.

A few commercial services currently offer specialized software for checking scientific images, but the programs aren't designed for large-scale, automated use. Ideally, a program would extract images from a scientific paper, then rapidly check them against a huge database, detecting copies or manipulations. Last year, several major scientific publishers, including Elsevier, Springer Nature, and EMBO Press, convened a working group to flesh out how editors might use such systems to pre-screen manuscripts. Efforts are under way—some funded by the O.R.I.—to create powerful machine-learning algorithms to do the job. But it's harder than one might think. Daniel Acuña, a computer scientist at Syracuse University, told me that such programs need to be trained on and tested against large data sets of published scientific images for which the "ground truth" is known: Doctored or not? A group in Berlin, funded by Elsevier, has been slowly building such a database, using images from retracted papers; some algorithm developers have also turned to Bik, who has shared her set of flawed papers with them.

Bik told me that she would welcome effective automated image-scanning systems, because they could find far more cases than she ever could. Still, even if an automated platform could identify problematic images, they would have to be reviewed by people. A computer can't recognize when research images have been duplicated for appropriate reasons, such as for reference purposes. And, if bad images are already in the published record, someone must hound journal editors or institutions until they take action. Around forty thousand papers have received comments on PubPeer, and, for the vast majority, "there's absolutely no response," Boris Barbour, a neuroscientist in Paris who is a volunteer organizer for PubPeer,

told me. “Even when somebody is clearly guilty of a career of cheating, it’s quite hard to see any justice done,” he said. “The scales are clearly tilted in the other direction.” Some journals are actively complicit in generating spurious papers; a former journal editor I spoke with described working at a highly profitable, low-tier publication that routinely accepted “unbelievably bad” manuscripts, which were riddled with plagiarism and blatantly faked images. Editors asked authors to supply alternative images, then published the studies after heavy editing. “I think what she’s showing is the tip of the iceberg,” the ex-editor said, of Bik.

Some university research-integrity officers point out, with chagrin, that whistleblowing about research misconduct on social media can tip off the scientists involved, allowing them to destroy evidence ahead of an investigation. But Bik and other watchdogs find that posting to social media creates more pressure for journals and institutions to respond. Some observers worry that the airing of dirty laundry risks undermining public faith in science. Bik believes that most research is trustworthy, and regards her work as a necessary part of science’s self-correcting mechanism; universities, she told me, may be loath to investigate faculty members who bring in grant money, and publishers may hesitate to retract bad articles, since every cited paper increases a journal’s citation ranking. (In recent years, some researchers have also sued journals over retractions.) She is appalled at how editors routinely accept weak excuses for image manipulation—it’s like “the dog ate my homework,” she said. Last year, she tweeted about a study in which she’d found more than ten problematic images; the researchers supplied substitute images, and the paper received a correction. “Ugh,” she wrote. “It is like finding doping in the urine of an athlete who just won the race, and then accepting a clean urine sample 2 weeks later.”

Last year, Bik’s friend Jon Cousins, a software entrepreneur, made a computer game called Dupesy, inspired by her work. One night, after Thai takeout, we

tried a beta version of the game at her computer. Bik's husband went first, clicking a link titled "Cat Faces."

A four-by-four panel of feline mugshots filled the screen. Some cats looked bug-eyed, others peeved. Instructions read, "Click the two unexpectedly similar images." Gerard easily spotted the duplicates in the first few rounds, then hit a more challenging panel and sighed.

"I see it, I see it," Bik sang quietly.

Finally, Gerard clicked the winning pair. He tried a few more Dupesy puzzle categories: a grid of rock-studded concrete walls, then "Coarse Fur," "London Map," and "Tokyo Buildings."

When my turn came, I started with "Coffee Beans." On one panel of dark-roasted beans, it took me thirty-one seconds to find the matching pair; on the next, six seconds. A few panels later, I was stuck. My eyes felt crossed. A nearby clock ticked loudly.

"Should I say when I see it?" Bik asked. "Or is that annoying?"

"No, no, no," Gerard said.

"Just tell me when it's annoying, because I don't always know," she said.

"Absolutely. You're annoying," he replied.

On her turn, Bik cruised swiftly through several rounds of "Coarse Fur," then checked out other puzzle links. Some panels were "much harder than my normal work," she said. The next day, Cousins e-mailed us with results: Bik's median time for solving the puzzles was twelve seconds, versus about twenty seconds for her husband and me.

A couple of weeks later, I called Jeremy Wilmer, a psychology researcher at Wellesley University, to ask what made Bik so good at picking out recurring visual motifs. Bik denies having a photographic memory, and says she's terrible at recognizing faces; people often assume that she must excel at "I Spy" and "Where's Waldo?," but she is good at discerning similarities, not differences. ("I cannot find Waldo," she joked.) Bik attributes her success to practice, and to being "crazy enough" to scrutinize images for hours on end. Wilmer, who co-directs TestMyBrain.org, a Web research project that administers standardized online tests of memory, perception, and cognition, agreed that the answer was probably some combination of genetics and accumulated expertise. His question was whether Bik's prowess reflected exceptional memory, perception, or both.

Wilmer set up seven tests online for Bik to take, including the gold-standard Cambridge Face Memory Test. On face recognition, she performed well below average. But she scored high on a task that called for memorizing fifty abstract art images and then picking out the ones she'd seen before. Her best performance was on a task in which the computer screen quickly alternated between two photos of real-world scenes. The photos were identical, except for one area, and separated, for a split-second, by a blank screen—an opportunity to forget what you've seen. Most people are lousy at detecting the differences. Bik scored in the ninety-ninth percentile. Wilmer told me that Bik is "incredibly good" at holding intricate scenes in her mind and comparing them. This talent had been useless to her, until it wasn't.

On the night we played Dupesy, Bik stayed up after Gerard went to bed, looking at images. Two side-by-side pictures of purple-stained cancer cells were the same, but one was rotated ninety degrees and had a few cells missing. "These people are not just rotators, they are also Photoshoppers," she said. She began posting her findings to PubPeer and updating her spreadsheet. The work had an immersive, repetitive rhythm. When I left, at around midnight, Bik was still working in silence under the collective gaze of her turtles. The next morning, I checked her

Twitter feed. “I am ringing the alarm,” she had tweeted, at 2 a.m., about a project she was pursuing with other, anonymous image sleuths; together, they had identified more than four hundred sketchy-looking articles that all appeared to come from the same source—a “paper mill,” somewhere in China, that sells faked English-language manuscripts.

Over the last year or so, Bik’s influence has grown. Not long after I visited her, a University of Maryland research group retracted or corrected several studies she had flagged. (“It’s always bittersweet,” Bik told me, of victory; such retractions are good for science, but can be a setback for an entire lab group.) Last March, on her blog, Bik dissected the French study that purported to show the effectiveness of hydroxychloroquine against COVID-19; among other issues, she noted that the manuscript had been vetted and accepted in just twenty-four hours, and that one of its co-authors was the editor-in-chief of the journal that had published it. (In a statement, the publishers said that, to minimize potential bias, manuscript review had been delegated to an associate editor.) A few weeks later, two other COVID-19 studies by other research groups were retracted from prominent medical journals. Examining the past work of one of the scientists involved in both reports, Bik found multiple, complex image duplications in photos of inner-ear tissue from rats, gerbils, and guinea pigs. Her findings were reported in BuzzFeed and other outlets.

Increasingly famous, Bik has become a target. Social-media trolls have attacked her, and critics have tried to spread misinformation about her. “There was an editing war on my Wikipedia page,” she told me. (Supporters corrected her entry.) Didier Raoult, the prominent microbiologist who led the hydroxychloroquine study—he heads the Institut Hospitalo-Universitaire Méditerranée Infection (I.H.U.), in Marseille—called her a “witchhunter” in a tweet; later, in a news interview on French television, he made reference to “une cinglée”—a crazy woman—who was criticizing his work. Bik posted PubPeer comments on more than sixty of his papers; Raoult’s colleague at the I.H.U., Eric Chabrière, “doxxed” her by

tweeting her residential address. Recently, he and Raoult filed a legal complaint in Marseille accusing Bik and Barbour of harassment, and Bik of attempted extortion (she rejects these allegations); two petitions in support of Bik have drawn thousands of signatures from scientists, and a science agency within the French government—the National Center for Scientific Research, where Barbour works—has issued a statement of support for Bik and Barbour, condemning Raoult and Chabrière’s “judicialization of scientific controversy and criticism.”

All the while, Bik has continued to rack up retractions. Last June, researchers at Harvard retracted a ten-month-old paper from *Nature*, and, in November, prominent Dutch scientists retracted a 2015 paper from *Science*; Bik had e-mailed the journals’ top editors about Western-blot duplications in both papers, and tweeted, tagging @Nature and @ScienceMagazine. (The authors maintained that their over-all findings had been confirmed.) Major publishers have also retracted several hundred faked paper-mill papers flagged by Bik and her sleuthing colleagues. The Microbiology Society, an organization in the U.K., has awarded her a prize for her work as a science communicator.

Recently, when I checked in with Bik, she told me that she was “a little bit stressed out.” She had been flooded with interview requests from journalists about the battle with the I.H.U. researchers. More and more universities and science organizations were asking her to give virtual talks. An introvert, Bik was somewhat overwhelmed. Video meetings and administrative busywork were consuming her life. Her in-box brimmed with unsolicited tips about papers to investigate. She missed the rhythmic work of perusing Western blots. “I feel I have no time left to do my regular work,” she said, wistfully. Not long ago, she found a window of opportunity. On Twitter, she posted four arrays of photographs of purple-stained cells. “#ImageForensics – Level advanced,” she tweeted. “Can you spot the overlapping panels?”