The One That Got Away?

A tale of modern science and its cruel twists, in which experienced physicists spend 10 years and millions of dollars chasing what most—but not all—now agree was a phantom

This is what the scientific discovery of a lifetime might look like: A decade ago, physicists sifting through the debris from collisions of heavy atomic nuclei observed a phenomenon they had never come across before, an event so pronounced they had virtually no doubt that what they were seeing was real. In the statistical parlance of science, the signal caused by this unknown phenomenon towered 6 standard deviations above the mundane background of known physics—enough to satisfy a 99.9999% confidence level that it wasn't a fluke. In the vernacular, it was a ticket to Stockholm.

Between 1983 and 1987, such peaks were discovered in two separate experiments at the Organization for Heavy-Ion Research (GSI) laboratory in Darmstadt, Germany. Both were operated by established, experienced physicists. A dozen experimental papers were published; a hundred more were filled with theoretical speculation that these puzzling signals might point to a particle unpredicted by conventional physics, or some bizarre event in the atomic nucleus.

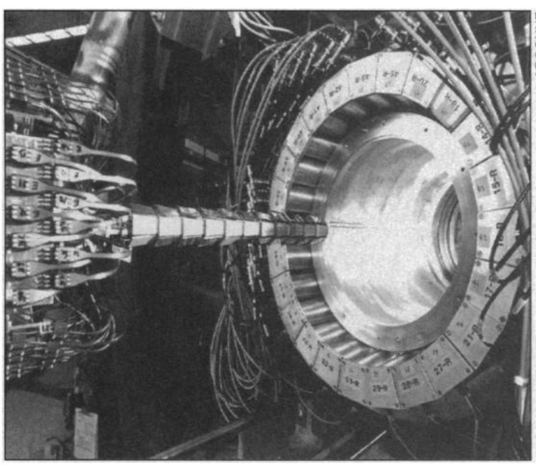
Now, a decade later, a new generation of experiments is reporting that this unknown phenomenon is nonexistent—not that it has a mundane explanation, but that it never existed in the first place. Some of the original discoverers, needless to say, are not happy. The 23 September 1996 issue of *Physical Review Letters* (*PRL*) ran a pair of argumentative comments. One, by physicists Jack Greenberg of Yale University and Tom Cowan of the Lawrence Livermore National Laboratory, challenged the way the new experiments were conducted and insisted that reports of the demise of the discovery are greatly exaggerated. The other, by physicists at the Argonne National Laboratory—where one of the new experiments, called APEX, was conducted—defended the negative conclusions.

The two comments are the manifestation of an acrimonious controversy that has been brewing for several years over the existence of this once-promising candidate for strange and unknown physics. On one side are Greenberg, Cowan, and Greenberg's colleagues from Yale—a group that includes a former President's Science Adviser who is president-elect of the American Physical Society, a former head of the High-Energy Physics Advisory Panel, and the present chair of the APS Division of Particles and Fields. They all continue to believe, or at least suspect, that the heavy-ion phenomenon is real. On the other side is virtually every nuclear physicist who has worked on the experiments, including some of Greenberg's former collaborators from the laboratory at GSI that first observed the effect.

While the immediate issue is a single physics question, the broader issues are pertinent to all of experimental science. This is a tale of the life and apparent death of a discovery, a case study in how experimental data should be interpreted, and an illustration of the questionable, perhaps delusive, power of statistics. And, finally, it is a story of how and why experiments end, and when researchers decide that it's time to give up—to admit that the original dramatic effect is irreproducible. As Michael Lubell, a City College of New York (CCNY) physicist who was involved in one negative search for the heavy-ion phenomenon, puts it, "For a 6-standard-deviation effect to disappear is very weird. There's no question about that."

Glimpse of a peak

The story begins in 1969, when the GSI laboratory began construction of an accelerator capable of smashing together nuclei as heavy as uranium. At the same time, University of Frankfurt theorist Walter Greiner pointed out that in such collisions, two heavy nuclei might touch momentarily just before they deflected off each other, forming a "quasi-atom" with an enormous electrical charge. Greiner speculated that the presence of such concentrated electric charge might lead to what Argonne physicist John Schiffer calls "strange and wonderful physics." In particular, the huge charge of the quasi-atom might serve to pull from the surrounding vacuum an electron paired with its antimatter counterpart—a positron. The postulated process was known as spontaneous positron emission because the electron would be swallowed up by the heavy ions and only the positron, the positively charged antimatter twin of the electron, would escape.



ARGONNE

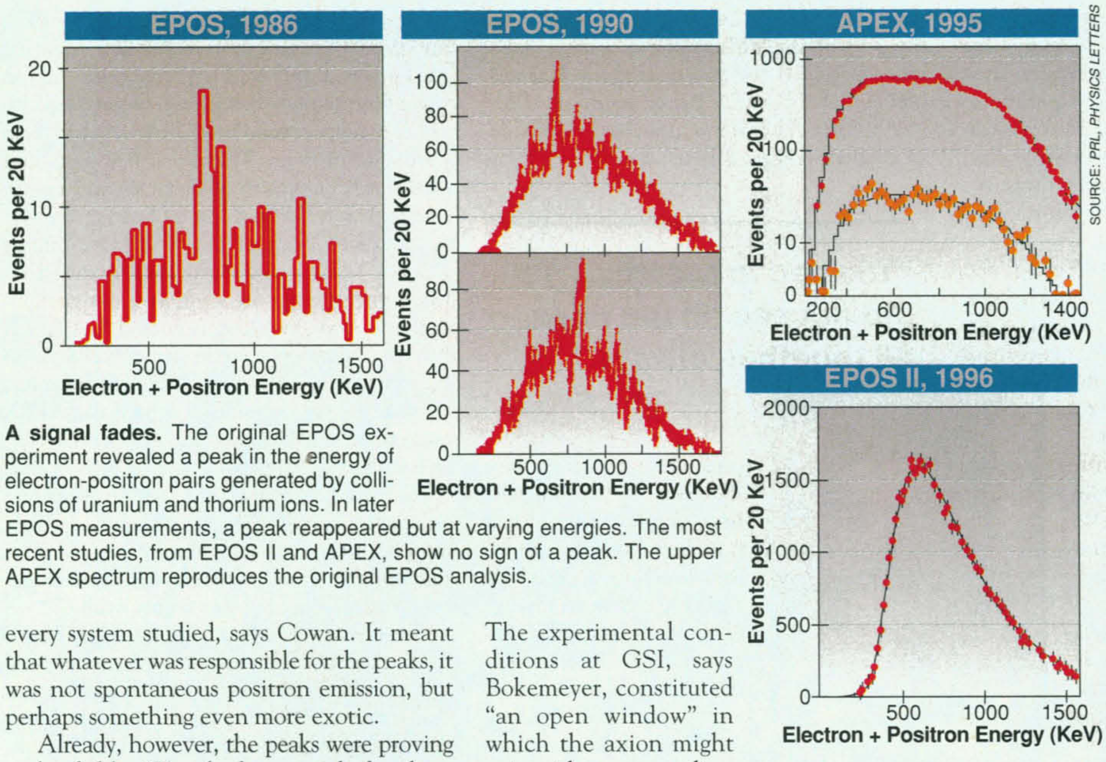
Dissecting a claim. A cylindrical sodium iodide array and a row of silicon detectors are components of APEX.

By 1976, the GSI accelerator had generated its first uranium beams, and 5 years later, physicists there were running experiments to look for the positrons. Two such experiments that would play major roles in the controversy to follow were known as ORANGE, led by Technical University of Munich physicist Paul Kienle, and EPOS, led by GSI's Helmut Bokemeyer and Greenberg, who was in Germany on sabbatical. The experiments slammed uranium ions into targets made of various heavy elements, creating a shower of electrons and photons and, once every million collisions or so, a single positron.

The key question was whether those exceedingly rare positrons were created by mundane physics, or by something more provocative, such as Greiner's proposed spontaneous positron emission. To tell the difference, the physicists generated an energy spectrum of the positrons—a plot of the number of positrons detected over a range of energies. Positrons created by routine physics would have a wide range of possible energies, resulting in a smooth spectrum. But spontaneous positron emission or some other exotic process might favor a single characteristic energy, creating a peak or hump in the spectrum.

In the very first experiments, which involved colliding uranium and curium, the EPOS physicists observed such a peak at 320 KeV (thousand electron volts). The peak matched theoretical predictions for spontaneous positron emission, but further experiments would tell for sure. Wrenching the positrons out of the vacuum would require the powerful electric field of a massive quasi-atom, which could only form in collisions of very heavy nuclei. If the peak came from spontaneous positron emission, says Cowan, who had joined the experiment as Greenberg's graduate student, "it would march down in energy and eventually disappear" as lighter and lighter nuclei were collided.

When the EPOS group looked at the debris from lighter nuclei, however, they saw no such pattern. "We saw the same peak or similar peaks" in



A signal fades. The original EPOS experiment revealed a peak in the energy of electron-positron pairs generated by collisions of uranium and thorium ions. In later EPOS measurements, a peak reappeared but at varying energies. The most recent studies, from EPOS II and APEX, show no sign of a peak. The upper APEX spectrum reproduces the original EPOS analysis.

every system studied, says Cowan. It meant that whatever was responsible for the peaks, it was not spontaneous positron emission, but perhaps something even more exotic.

Already, however, the peaks were proving to be fickle. "You had to search for these things," says Cowan. "[And] you did not see all the lines, all together, always." Greenberg and his collaborators thought the phenomenon responsible for the peaks might be extremely sensitive to the circumstances of the collisions. They came to believe that the accelerator had to be carefully tuned to the right energy, which wasn't easy at GSI, and that the target had to be precisely the right thickness. At least some of their doubts were eased, however, by the ORANGE collaboration, which was reporting peaks as well—different peaks, says Kienle, but then again, they were running their experiments at slightly different conditions.

With peaks appearing for various pairs of projectile and target nuclei, as well as in two different detectors, says University of California, Berkeley, physicist Stuart Freedman, an APEX collaborator, the outside world began to take them seriously. The most intriguing theoretical scenario—which Greenberg was the first to suggest, albeit cautiously—was that the peaks consisted of positrons from the decay of some other object created in the collisions. The two ions kissed, so the theory went, and in the huge electric field of the quasi-atom, a new and hitherto unknown neutral particle was created that then decayed into a positron and an electron.

"Fortunately or unfortunately," says Bokemeyer, the GSI reports came out at a time when theorists were speculating about a possible heavy neutral particle known as the axion, which might explain, among other things, the missing mass in the universe as well as some quirks in the theory of the strong force that binds the atomic nucleus.

The experimental conditions at GSI, says Bokemeyer, constituted "an open window" in which the axion might exist without contradicting other experiments that hadn't seen it. And the unambiguous discovery of any new particle, axion or not, would be a Nobel-caliber breakthrough.

Shifty signals

Now, the two experiments' participants went particle hunting. Both groups refit their apparatus to identify the electron that should be emitted along with the positron when the unknown particle decayed. Starting in 1985, EPOS looked at collisions of uranium and thorium and saw a 6-standard-deviation peak at 760 KeV (the combined energy of the electron and positron). The following year, however, the collaboration repeated the experiment and the 760-KeV line mysteriously vanished. "This was never understood," says EPOS physicist Piotr Salabura, now with the Institute for Nuclear Physics in Kraków, Poland. Instead, EPOS observed two more peaks of nearly 5 standard deviations, at 610 and 810 KeV, along with other peaks at similar energies when they collided uranium and tantalum.

These peaks seemed to be reproducible, but in a typically unsatisfying way. They would come and go for no apparent reason, a behavior that Greenberg and his colleagues again attributed to subtle changes in the energy of the colliding nuclei. Greenberg was so concerned about the apparent energy-dependence of the peaks, he says, that he would continuously ride the GSI machine operators to make sure that the beam energy was precisely what the physicists had specified. Absolute precision was crucial, Greenberg recalls. Once, for instance, the researchers saw the 810-KeV peak in the data; when they went back the next day at what they thought was the same energy, the peak did not reappear. "We said it must be a fluke," says Greenberg. "Let's make sure the energy was right. The chief operator looks it up, and it was wrong. He retools the machine to the energy where it was before, and the peak reappears."

The ORANGE collaboration, meanwhile, was more or less reproducing the results—with the "more or less" depending on precisely how you defined reproduction. After the EPOS physicists showed their colleagues and competitors in ORANGE a 625-KeV peak seen in uranium-thorium collisions, the ORANGE physicists, who were on the verge of giving up, decided to reproduce precisely the experimental conditions that had yielded that particular peak at EPOS. When they did so, they saw a peak at 635 KeV, which seemed close enough, given the experimental uncertainty. And this peak was huge—"something like a 6.5-standard-deviation effect," says collaboration member Wolfgang Koenig. As Salabura puts it, "[Now,] we believed the phenomenon was reproducible."

The reviews from the physics community, however, were still decidedly mixed. Some, like physicist Lawrence Krauss of Case Western Reserve University in Cleveland, who was then at Yale, considered the results "kosher" and spent much of their time trying to hone the theoretical explanation. Others, such as Berndt Muller, a Duke University physicist who was working with Greiner at the time, began to suspect that the GSI physicists might be unknowingly propagating "pathological science."

This was a term coined in the 1950s by Nobel laureate chemist Irving Langmuir to describe research in which, as Langmuir put it, "people are tricked into false results by a lack of understanding about what human beings can do to themselves in the way of being led astray by subjective effects, wishful thinking, or threshold interactions." At a 1986 nuclear physics conference, says Muller, he went through Langmuir's criteria point by point, suggesting they fit the heavy-ion experiments all too well. "It created a tremendous amount of animosity," he recalls.

A vanishing act

In 1989, GSI suspended its experimental program while it rebuilt its heavy-ion synchrotron.

The EPOS and ORANGE researchers dispersed to pursue the mysterious peaks in other venues in Europe and the United States. Their approach would be to run the experiments backward by a technique called Bhabha scattering: colliding a beam of positrons into electrons in hopes of creating whatever it was that might be decaying into the electron-positron pairs in the heavy-ion collisions.

Greenberg joined Mike Lubell of CCNY, Kelvin Lynn of Brookhaven National Laboratory, and a handful of graduate students, visiting scholars, and postdocs to try the experiment at Brookhaven. There, Lynn and the students spent a year refitting a mothballed electron accelerator to produce a beam of positrons with an energy that could be carefully controlled. They then took data for nearly two more years, says Lynn, "nonstop, 24 hours a day ... including Christmas."

But conflict arose between Greenberg and the other physicists. By the time the experiment was over, all but one of Lynn's graduate students and two of Greenberg's had abandoned it. Lynn describes the experience as "the most difficult interaction" of his career. (Lynn contacted *Science* at Greenberg's request.) Lynn, Lubell, and the graduate students all say the problem was Greenberg's insistence that the experiment be done as thoroughly and cautiously as humanly possible, combined with an unshakable belief that the peaks at GSI had been real, and that they would appear in the Bhabha-scattering experiment if it was done correctly. He was "absolutely convinced," says Lubell. Adds Lynn, "Jack was on this Nobel Prize hunt. ... Jack was so convinced from his GSI data that it had to be there, it was like a religion with him."

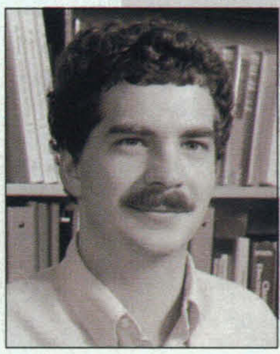
Nonetheless, after the nearly 2 years of data collection and two more years of analysis, Lynn and Greenberg finally published a pair of papers saying that they had seen nothing. While Muller calls the experiment and the final analysis "brilliant," Lynn adds that Greenberg was not satisfied with the negative result and would have preferred to keep going. Lynn says that he and Peter Bond, head of the Brookhaven physics department, had to "[pull] the plug on the experiment."

The other Bhabha-scattering experiments came up empty as well. But theorists offered plenty of reasons why a mysterious effect could manifest itself in the heavy-ion experiments but not in Bhabha scattering. And three new experiments designed to gather orders of magnitude more data were already in the works to settle the issue definitively.

Starting in the late 1980s, the remaining EPOS physicists spent 5 years building a completely new experimental apparatus to study the peaks; this became the EPOS II collaboration. The ORANGE collaboration refitted their apparatus. And a multiuniversity collaboration designed and constructed the $2.3 million ATLAS Positron Experiment, or APEX, at Argonne. The motivation, says APEX cospokesperson Russell Betts, was simple: If the peaks were real, they were "fantastic, something very new, very exciting. But it was clear the early [results] were kind of marginal." By 1993, APEX, with Greenberg as one of the collaborators, EPOS II, and ORANGE were all taking data.

"You had to search. ... You did not see all the lines, all together, always."

—Tom Cowan



In October 1995, the APEX physicists were the first to publish their results, in *PRL*. They had been unable to reproduce the GSI peaks. As Argonne's Schiffer puts it, "We found absolutely no sign of the [peaks] that had been reported before."

Although Greenberg was still a member of the collaboration, his name was not on the paper. He says he "agonized" over the decision to withdraw it, but he believed that APEX had neither reproduced the conditions of the original EPOS experiment, nor run long enough, nor systematically collected enough data to make such a definitive statement. He believed that the collaborators had cut the experiments short to beat out the competition at GSI. Moreover, Greenberg and his postdoc Guangsheng Xu had done their own analysis of the APEX data and found a peak—"very reminiscent of the peak we saw at [the original] EPOS," says Greenberg.

At a meeting in Bloomington, Indiana, the same month the APEX paper came out, says Bond, Greenberg and the APEX physicists had a heated argument over the existence or nonexistence of the peak and the validity of the APEX analysis. "The collaborators were so mad at Jack they could hardly speak," says Bond, "and Jack was furious at them." Yale physicist Allan Bromley, president-elect of the APS and a former White House science adviser, describes the dispute simply as "a very difficult and unpleasant business. No question about it."

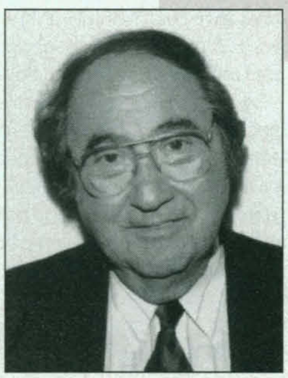
With Cowan, Greenberg submitted his criticisms of the APEX experiment to *PRL*, and the APEX physicists responded, resulting in the thrust and counterthrust published in the 23 September 1996 issue. Greenberg also enlisted support from his colleagues at Yale, in particular Bromley; Michael Zeller, past chair of the APS Division of Particles and Fields; and Jack Sandweiss, a former head of the Department of Energy's (DOE's) High-Energy Physics Advisory Panel. All three believe that Greenberg's peaks are suggestive of a real, albeit unknown, phenomenon, and that APEX has an obligation to redo the experiment to track it down. "The problem," says Bromley, "is that the experiment for which the APEX equipment was designed has not been done." It hasn't tried the specific thin targets that EPOS used, nor has it carefully scanned through a range of beam energies to truly search for the elusive peaks, he says. "And it may well be shut down before [it] gives a definitive answer. That just seems wrong."

Bromley even called DOE to plead Greenberg's case, but the DOE administrators sided with the APEX physicists, who believe that the case is closed. "We have funded good people to do this experiment," says David Hendrie, director of the DOE division of nuclear physics. "We have to acquiesce with what they have found. I personally believe the phenomenon has gone away. At some point, Jack must have smelled Nobel Prize in this stuff. He has been very reluctant to write it off."

The new results from the two other experiments, ORANGE and EPOS II, didn't help Greenberg's case. Both detectors had nearly 10 times the sensitivity of the original versions, and they had collected an order of magnitude more data. They had precisely reproduced the conditions of the early experiments—"we made sure those targets were thin enough, and so on," says Bokemeyer—and done what Greenberg was insisting APEX do, which was to increase the energy of the beam slowly in very small increments, one step at a time, looking for just the right energy to create a peak. And although Greenberg believes there is evidence of a peak in the EPOS II data, the researchers involved in both experiments say they saw nothing.

"With any kind of new phenomenon, you try to see what conditions would optimize it."

—Jack Greenberg



But how could the peaks have vanished? Or rather, why had they been seen in the first place? Last June, physicists from all three experiments met in Oxford, U.K., to discuss what had happened and decide what to do next. To those who hadn't been on the original round of experiments, the meeting was a revelation.

Unkind cuts

At issue was how to analyze a potential new phenomenon, whose signal can barely be detected in a sea of background noise. It's the essence of experimental physics, says Bokemeyer. With a few celebrated exceptions, "any new effect will start as a tiny signal indistinguishable at first from the background." To pull a signal out of the noise, researchers have to apply "cuts" or "gates" to the data. Ideally, these cuts have plausible physics behind them. For instance, if the GSI peaks came from the decay of a neutral object, then the electrons and positrons should have been emitted in opposite directions to conserve momentum. Thus, the physicists could look specifically for those back-to-back electrons and positrons as they collected and analyzed the data, and reject uncorrelated pairs.

But making such cuts can be a dangerous game. Unphysical cuts can turn a meaningless statistical fluctuation into a 5- or 6-standard-deviation peak. The danger grows when the phenomenon is both tantalizing and completely unexpected. "You should always be guided by some physical scenario," says Koenig, "but here there were hundreds of papers discussing various scenarios, some very flexible." Moreover, the EPOS physicists felt pressured to identify fruitful experimental conditions, because their data came slowly—a few "counts" of possible signal per day. This effort to optimize the data collection led to decisions that other physicists now find questionable.

Take what the EPOS physicists referred to as the top-hat criterion. Bokemeyer says that the EPOS physicists had noticed that what turned out to be peaks in the final analysis would first appear online as a top-hat–shaped bulge in an otherwise smooth spectrum. So the experimenters would start collecting data at a particular energy or with a particular target, and if the spectra were smooth and flat, they would stop the experiment. "We would change the energy or target and try again," says Bokemeyer. "When the spectra started to look like a top-hat, this seemed to be the correct [conditions], and we would continue running without interruption."

Greenberg argues that such practices are the only way to pursue such tenuous phenomena. "With any kind of new phenomenon, you try to see what conditions would optimize it," he told *Science*. "Obviously, not having any theoretical guidance, we tried various empirical things that would optimize the appearance of this peak." But to some of the APEX physicists, all this seemed like a sure way to force a statistical fluctuation into a peak, just as you can "prove" that a coin preferentially lands on heads in runs of 100 flips by stopping any run if it doesn't show an early preponderance of heads after, say, 30 flips. "You could always find a peak," says physicist Frank Wolfs of the University of Rochester in New York, a co-spokesperson for APEX, "because you could always find reasons to reject one data set when you did not see it, for example by saying the target quality was not good for that particular run."

"The problem is you can't prove something isn't there."

—Russell Betts

The only way to prove that peaks found in such analyses are real, explains Bokemeyer, is to repeat the experiment and use the identical analysis and cuts to see if a peak appears out of an entirely new set of data. If the peak is illusory—a product of statistical fluctuations combined with unphysical cuts—it's unlikely to reappear in the new data set. Indeed, when the EPOS physicists tried to reproduce their peaks, they rarely reappeared twice in a row at the same energy. But Greenberg and his collaborators didn't see that as fatal. With no solid physics to explain the peaks, explains physicist Rudi Ganz, who wrote his thesis on EPOS II and is now at the University of Illinois, Chicago, any peak that did appear could always be considered a possible confirmation of some previous peak at a nearby energy.

At Oxford, however, Ganz presented an analysis showing how such practices could create huge peaks out of meaningless statistical fluctuations in the data. After EPOS II had seen one huge peak come and then go in two apparently identical experiments, explains Ganz, he decided to try an unusual analysis on the next run. In a demonstration that APEX collaborator Freedman, for one, calls "really extraordinary," Ganz began by creating a random-number generator that assigned to each collision a number between 0 and 1. He then divided the data into two identical sets: "Everything above 0.5 is data set one," he says, "and everything below that is data set two." Ganz then took the first data set and started looking for peaks, using cuts similar to those in the original EPOS experiments—for instance, defining the time of flight of the electrons and positrons from the collision to the detectors.

"You have really powerful computers," he says. "So you just sit down, make a cut, and you get a spectrum in a few seconds. You say the time of flight should be between 8 to 20 nanoseconds. You get the spectrum and find nothing. So you take the data from 0 to 8 nanoseconds. Now, you see at one point in the spectrum a little structure. So you go from 7 to 8 nanoseconds, and you can really improve the signal by playing around with the borders and adjusting what is optimum. Because you really have no idea what you're looking for, you have the freedom to choose any time of flight cuts."

Using such liberal cuts, Ganz managed to find an enormous peak in the first data set at 655 KeV. Yet, when he applied the identical analysis and cuts to the twin set, there was nothing. It was confirmation, says Ganz, "that these structures, even if they look nice, aren't necessarily reproducible or statistically significant." Greenberg regards Ganz's analysis as a meaningless caricature of the procedures used at EPOS. But other EPOS physicists disagree. And even Bromley says, "If it is as represented, it is a very bad sign."

By the end of the Oxford meeting, none of the physicists involved with the latest experiments seemed to see a reason to continue looking for the peaks. "I think we all overestimated the statistical relevance of the peaks we saw," says EPOS II co-spokesperson Dirk Schwalm, of the Max Planck Institute for Nuclear Physics in Heidelberg, Germany. "It sounds a bit silly in the end, 10 years later, but I think that's what happened." As Bokemeyer put it, they had taken a questionable situation in physics and clarified it.

As for the peak Greenberg still insists can be found in the APEX data, few physicists other than Cowan and Greenberg's colleagues at Yale seem to give it much credence. The EPOS II physicists—including Salabura, who was first author on the last EPOS paper reporting the existence of the peaks—say if it were real, they would have seen it. And the APEX physicists don't think it is anything more than yet another, albeit considerably smaller, statistical fluctuation. Says Betts, "Whatever it is he's seeing, it ain't physics."

Greenberg wants just one more APEX experiment to try to reproduce his latest peak. "If they don't find it, I'll quit looking for it," he says. But the APEX physicists are having none of it. "The problem is you can't prove something isn't there," says Betts. "At the end, it's somebody's considered scientific judgment that it's no longer worth their time. We've all put a lot of time and effort into this; a lot of grief has gone into this. We all feel no, enough. No more."

–**Gary Taubes**