Statisticians at Odds Over Software Ownership

A nasty fight has erupted over similarities—and alleged errors—in two software packages widely used by academics

THE LINE BETWEEN ACADEMIC AND BUSINESS interests is getting blurred in a dispute among statisticians that raged through the computer networks in the form of an E-mail debate in December. It centers on a charge of plagiarism made by Leland Wilkinson, a well-known statistician at Northwestern University in Illinois, against another professor—psychologist Pawel Lewicki of the University of Tulsa, in Oklahoma.

In addition to their academic posts, Wilkinson and Lewicki head competing companies that market statistical software packages widely used by academic researchers to draw graphs from complex data sets. Wilkinson, whose company, SYSTAT, is the older of the two, claims that Lewicki copied key features of SYSTAT's programs and put them into a competing product that Lewicki's company, StatSoft, is peddling to Wilkinson's own customers. Wilkinson also claims that Lewicki's software is conceptually flawed. Lewicki claims the same about Wilkinson's package, while flatly denying the charges against himself. Lewicki says Wilkinson has made "unforgivable mistakes" in a 26-page pamphlet he mailed out to 40,000 of the nation's statisticians setting forth his charges against StatSoft.

The case illustrates how difficult it can be to resolve intellectual property disputes, particularly when claimants are fighting over the use of algorithms in software. (Usually, all that inventors can protect is the "look and feel" of a program as expressed on a computer screen.) The SYSTAT-StatSoft row also suggests how hard it can be to disentangle business from intellectual interests. For example, another well-known statistics professor contacted by *Science* declined to go public with his comments, fearing he might provoke a barrage of negative advertising against his own small software company. And beyond the quarrel over priority and plagiarism is a broad debate about what can be done to raise the quality of all statistical software programs.

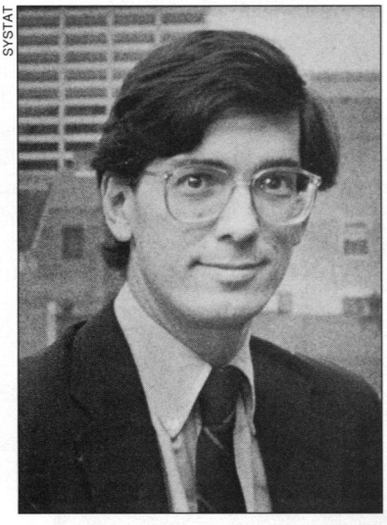
Although both men say they wish to avoid going to court, Lewicki says he has asked his lawyers to prepare a lawsuit. Meanwhile, statisticians—alerted to the fracas by Wilkinson's vast mailing and an equally detailed response from Lewicki and StatSoft—have now begun their own debate in a forum known as "Stat-L," an electronic bulletin board. Many of them seem torn between a desire to respond to the "cry from a wounded scholar" (as Canadian psychologist John Vokey wrote of the dispute) and a reluctance to get dragged into a brawl. For example, one statistician urged his colleagues to stay out of the fight, because it involves "plagiarism, copyrights, etc.—money, not statistics." But he seems to be in the minority. Most who chipped in on the Stat-L discussion thought the profession should police itself through public debate, and that the issues in this case—especially the question of whether statistics software packages can be trusted—are too significant to be left to the courts.

There does appear to be one point of agreement between the disputants: It was StatSoft's aggressive sales campaign and its growing success in the marketplace early this year that set the dry tinder ablaze. StatSoft sent out promotional material last year claiming that its statistical program is faster and less "tedious" to use than SYSTAT's—which Lewicki insists is the basis for Wilkinson's dismay. Wilkinson, who says he had objected privately to Lewicki for several years, finally decided to mail out his counterattack ("The Truth about StatSoft") in October because people were becoming "confused" about the identity of the two products and were buying more and more copies of StatSoft's package.

In his broadside, Wilkinson traces the dispute all the way back to 1984. In April of that year, he claims, Lewicki telephoned him from Tulsa and asked for a free "review" copy of SYSTAT for an article to appear in a professional journal. Wilkinson obliged. The article never appeared, but 3 years later, Lewicki released his own statistics program called CSS. Wilkinson alleges that many of the procedures in CSS were "identical to or close imitations of" those in SYSTAT. Lewicki's new statistical graphics package, CSS:Statistica, came out in 1991. Wilkinson says in his pamphlet that it was "copied" substantially from SYSTAT.

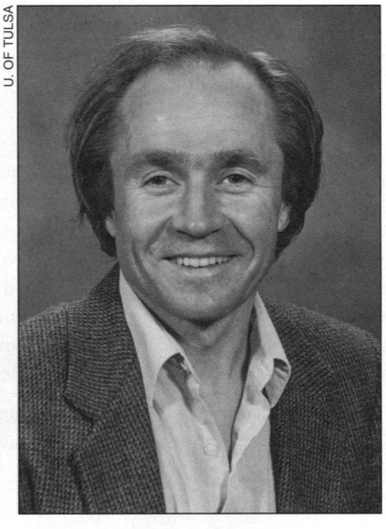
Lewicki has now mailed out a 48-page booklet in response to Wilkinson's, charging that Wilkinson's remarks are "entirely false." Lewicki says he did receive a review copy of SYSTAT and use it in writing a review, but that the article was rejected by the editors of *American Psychologist*. Besides, what difference does that make? he asks. "If we wanted to reverse engineer [the software] we could have purchased a copy." The allegation that Lewicki plagiarized copyright material verges on "criminal wrongdoing," Lewicki says. In the case of one graphic tool in particular—a "cluster graph"—which Wilkinson claims to have derived from an idea of a colleague at Yale, Lewicki scoffs that it "was not only published before but included in statistical packages before SYSTAT was even started. It's hard to imagine where he got those ideas." Indeed, the booklet insists that far from being the model for StatSoft's programs, SYSTAT is a less capable package. Lewicki gives numerous examples of test

SYSTAT®



SYSTAT

StatSoft™



U. OF TULSA

Hardball on software. *Statistician Leland Wilkinson* (above) *accuses Pawel Lewicki of copying his software.*

data files that he claims are not handled correctly or conveniently by SYSTAT.

Wilkinson's charges and Lewicki's countercharges have struck a nerve among academics. According to William Eddy, a statistics professor at Carnegie-Mellon University in Pittsburgh and chairman of the National Research Council's (NRC) Committee on Applied and Theoretical Statistics, leaders in the field have been trying for 15 years to establish guidelines for computer programs, and it has been hard to reach a consensus, because every program has its flaws and its devotees. "Every package has a bug," says Paul Tukey of Bellcore. "It's a threshold question" as to how many and what kind are acceptable. The NRC panel held a meeting on the subject earlier this year and will be issuing papers soon, Eddy says.

A key point of contention is the ability of the competing programs to handle "ill-balanced" data sets. For example, Wilkinson and Lewicki both say their adversary's program yields poor results in an Analysis of Variants, or ANOVA, procedure, often used to compare results from experiments in which some subjects receive a treatment and other "controls" receive none. The problem is that if this procedure is done mechanically, it will yield some "cells" that are devoid of data and should not be used to generate additional data. Wilkinson claims that his adversary's program merely warns that some results "may not be estimated correctly because of missing cells," and then goes on to print erroneous data, while his own program, in contrast, grinds to a halt and asks the user to try another analytical method.

In defending against these allegations, StatSoft presented a long discussion of flaws that allegedly appear in SYSTAT when there are gaps in the data, citing a statistics textbook coauthored by University of Kansas professor Dallas Johnson. But Johnson himself isn't buying StatSoft's contentions. Wilkinson sent Johnson a copy of a page from StatSoft's pamphlet. Johnson's opinion: Much of it "has no meaning at all." The procedure used by StatSoft as an illustration "shouldn't be computed at all," says Johnson. His own view is that "it's better not to produce than to produce something wrong," and he says he favors programs that balk at erroneous instructions.

The debate rages on, having now entered the rebuttal-of-rebuttals stage with no end in sight. Despite the anguish it may be causing the two protagonists, however, it may be doing some good. Already, both sides have identified flaws in the other's program, and this will lead to quick improvements. And potential users are getting a graphic lesson in the limitations of statistical software packages. ■ ELIOT MARSHALL

A Lethal "Cold Fusion" Blast

In the latest, deadly chapter in the bizarre history of "cold fusion," an explosion of a deuterium-palladium electrolysis experiment at SRI International in Menlo Park, California, killed one researcher and injured three others on 2 January. The fatally wounded scientist, Andrew Riley, was a contract researcher for the Electric Power Research Institute (EPRI), which has funded SRI's research on deuterium-palladium electrolysis since 1989. Also injured were laboratory director Michael McKubre, SRI researcher Stuart Smedley, and EPRI contract scientist Steven Crouch-Baker.

Few details of the explosion were available as *Science* went to press. SRI spokesman Dennis Maxwell was quoted in the *San Francisco Chronicle* as saying the accident occurred while three of the scientists were lifting a steel cylinder containing a palladium electrode in a deuterium oxide electrolyte from a water bath and placing it on a shelf. The *Chronicle* also reported that an emergency services officer said the explosion took place when one of the scientists attempted to open a jammed valve on a cylinder containing deuterium. Maxwell could not be reached for comment, but an SRI spokeswoman confirmed his statement.

At the Second International Cold Fusion Conference in Como, Italy, last June, McKubre reported measuring reproducible "excess" heat in a deuterium-palladium electrolysis cell. Although neither McKubre nor members of his laboratory returned calls from *Science*, EPRI program manager Joseph Santucci says that McKubre had achieved reproducible excess heat production by discovering how to "load" the palladium electrodes with deuterium molecules at an atomic ratio approaching unity. The explosion occurred during an attempt to "scale up" McKubre's earlier experiments, Santucci said.

Hydrogen explosions in electrolysis experiments involving palladium are not unknown, since the metal can catalyze an explosive recombination of hydrogen and oxygen. Santucci admits that such a conventional explanation might account for the explosion, but he claims it is "unlikely," since preliminary information suggested that the energies released were "substantial." More information will come to light over the next 2 or 3 weeks, as EPRI investigators pore over the cell's remains, analyze the palladium electrode, and take readings from some 17 instruments that were recording data when the cell exploded. ■ DAVID P. HAMILTON

Russian Academy: So Far, So Good

Russia's beleaguered scientists finally have a little good news to celebrate: Last month's forced merger of the All-Union Academy of Sciences (the key scientific institution of the Soviet Union for more than 50 years) with its just-created rival, the Russian Academy of Sciences, has so far gone much more smoothly than expected. Instead of dismantling the academy and flooding it with political appointees—as many top scientists feared—applied mathematician Yuri Osipov, the newly elected president of the combined Russian Academy of Sciences, is instead protecting academy scientists and bringing in democratic reform.

Last November, when it became plain that the All-Union Academy would be forced to merge with the Russian Academy, many scientists complained that control of research might rapidly pass out of the hands of the nation's best scientists. Biochemist Evgenii Sverdlov, a corresponding member of the old All-Union Academy, was among those who feared the worst (*Science*, 20 December 1991, p. 1717). But now Sverdlov has changed his tune. "The situation in the academy has improved dramatically," Sverdlov told *Science* from Moscow this week. He says that the few new members who have been brought into the academy are all reputable scientists.

And Osipov—an early supporter of the rival academy—has turned out to be "more or less the right person" to run the combined academy, says Sverdlov. Thanks to his close links to Boris Yeltsin (both are from Sverdlovsk) he has been able to argue for increased support for science. New democratic institutions, including the free press and an elected parliament, have also enabled scientists to put pressure on the Yeltsin government, says Sverdlov. One result: Yeltsin appeared before the inaugural session of the merged academy and assured scientists that the Russian government would double their salaries.

But daunting problems remain. Present levels of funding for the academy's institutes will continue only until March; then decisions will have to be made on cutting staff and closing unproductive institutes. That is when a mighty power struggle will begin as scientists fight to keep their own institutes—and jobs—alive. ■ STEVEN DICKMAN