Is the *p*-Form of the *t*-Table Fraudulent?
Yes, it Misrepresents Scientific Thought (and Credit)[1]

by Stephen T. Ziliak

Professor of Economics
College of Arts and Sciences
Roosevelt University
430 S. Michigan Ave
Chicago, IL 60605 U.S.A.
Email: sziliak@roosevelt.edu
Website: http://faculty.roosevelt.edu/Ziliak

October 28, 2008

[1] With apologies to Peter Medawar, for borrowing the form of the title to his article, "Is the Scientific Paper Fraudulent? Yes, it Misrepresents Scientific Thought," *Saturday Review*,, 1 August, 1964: 42-43.

Is the *p*-Form of the *t*-Table Fraudulent?

By Stephen T. Ziliak

Abstract Forms of arguments persuade scientists only insofar as those same scientists are free to choose to ignore them. When in 1925 Ronald A. Fisher (1890-1962) copyrighted and published "Student's" *t*-table in his own name, Fisher at the same time presented "Student's" table in a new *form*, emphasizing *p*-values. In 1938, Fisher again undertook to transfer the copyright on "Student's" table to him and his publishing house, Oliver & Boyd, but this time, without crediting "Student's" work, inventions, and previously copyrighted tables, 1908 to 1925—at all. Against "Student's" (1876-1937) warnings about the illogic and dangers of using *p*, Fisher used the *p*-form of "Student's" table to invent and spread worldwide an arbitrary rule about statistical significance – the 5 percent rule. In the life and social sciences, Fisher's *p* and rule continue to dominate journal editorial policy and science itself whenever and wherever a question of scientific inference has been posed. This article shows how such a quiet revolution and theft could occur. Acquiring copyrights to "Student's" table was probably not necessary for winning intellectual rights to "Student's" test. But it may have been sufficient.

Keywords: evidence, testing, interpretation, uncertainty, unity of science.

Is the *p*-Form of the *t*-Table Fraudulent?

By Stephen T. Ziliak

> There remains no consensus on whether to use *P*, *p*, P, or p values
> D. G. Altman 1991, p. 1902

Leading journals and patrons of science continue to enforce illogical standards of statistical "significance," judged by what now appears to be a fraudulent *t*-table.

As a matter of fact the *p*-value—the probability of observing a value of "Student's" *t* in excess of the value actually observed, given the sample size and experimental conditions—plays a contradictory role in logics of scientific discovery and justification. In the theatre of science after Ronald A. Fisher (1890-1962), *p*'s role is simultaneously that of the dominating leading man and forgettable extra.

Evidently the *p*-value's job is to sanctify (in fields such as medicine, biology, and economics) the final scientific output, with conclusive, George Clooney-like demonstrations of "statistical" significance ($p < .05$), or of the lack of it, "insignificance" ($p \geq .05$). The great mathematical statistician, Ronald A. Fisher, said: "it is principally by the aid of [significance testing, using "Student's" methods] that these studies may be raised to the rank of sciences" (Fisher 1925, p. 2). As Fisher, a great leading man of science, would have it, Fisher's *p* value became the center of attention, a proof of rank, in the newly quantitative sciences, where it remains (**** and

***** 2008, chps. 20-23). Meantime, philosophers and historians of social science, such as Gerd Gigerenzer, et al., (1989) and Donald MacKenzie (1981), having considered at length the history and logic of *p*, argue that Fisher's "confounded *p*-value" (Lang, Rothman, and Kahn 1998; **** and *****, p. 2) takes attention away from what should be in sciences after Galton the main event: estimation and interpretation of experimental magnitudes and control of real error, that is, control of random but especially non-random error (see also Reiss 2007).

A *p*-value, philosophers and scientists agree, does not decide what is scientifically important, and cannot (***** and ***** 1996; ***** and ***** 2008). Thus it takes more than a cult-like faith in the objectivity of "statistical" significance for a publishing scientist or historian to use a *p* value to decide an experiment, ban a drug, believe an economic hypothesis, or worship a journal's "impact factor," using *p* as the guide. Yet leading journals of life and social science, from the *American Economic Review* to the *British Medical Journal*, suggest that scientists over the past century have had a large non-random appetite for Fisher's *p* values.[1]

After Fisher's two fat and influential books of the 1920s and 1930s, and especially after the 1970s and the coming of computers, it is not easy to identify which (if any) part of the appetite for *p*-values exists by choice. For example, scientist-authors continue to be forced by journal editorial policy and referees to report "significance" if *p* is less than some uniform and pre-defined level, .05, for example, or .01. Authors are forced to shove aside—if not drop altogether—those variates and models which fail to cross the contrived line, no matter the expected effect of a decision to "accept" the null hypothesis of "no experimental difference" in the shoved-aside material (J. Freiman, et al., 1978). "Insignificance," measured by the *p*-value, is always and everywhere supposed to signify "unimportance," measured by the judgments of a community of scientists. This is one example of how the cult of statistical significance operates in science.

Of course some of us have more than passing knowledge of the cult and its strange-sounding rules and rituals. And what we find in investigations of use of *p* values after the meteoric rise to authority of Ronald A. Fisher (1890-1962) is an autocrat to whom today's compulsive use of *p* can be originally traced (Fisher 1925a, 1926, 1935, 1956). Students of "statistical significance," from "Student" himself to Ian Hacking and many others, find that restrictive policy rules about the use of *p* values in science are grinding real science to a halt, one regression at a time. (For a fuller list of names of distinguished critics of *p,* see page 2 of ***** and ***** 2008. See also Gigerenzer, et al., 1989, for a fine sociological history of Fisher's great influence on statistical practice.) As Arnold Zellner, a past-president of the American Statistical Association, wrote long ago: "The rationale for the 5% 'accept-reject syndrome' . . . requires immediate attention" (Zellner 1984, p. 277). That was 25 years ago. Still, use of *p* rages on—and how. For instance, the biostatistician Steven Goodman (1999a, p. 995), speaking of the "*p*-value fallacy," observed ruefully in *Annals of Internal Medicine*: "biological understanding and previous research play little formal role in the interpretation of quantitative results."

Regardless of what one means by "immediate attention" and "interpretation of quantitative results," new historical evidence on the *p*-form of the *t*-table suggests that biologists, economists, and other testing scientists —including some historians—are now standing face-to-face with what Nietzsche called "the twilight of the idols": the *t*-table, in its *p*-form, and judged by a *p*-rule, is not merely illogical and scientifically inconclusive—it was stolen.

This article uses previously neglected statistical tables, articles, and letter correspondence, by and between William Sealy Gosset (1876-1937) aka "Student" and "Student's" younger friend and colleague, Ronald A. Fisher (1890-1962), to reveal the intellectual and copyright history of "Student's" table, now used daily in science to make or unmake a legitimate scientific claim. The history of Fisher's quiet theft of "Student's" *t* table, 1925 to 1938, and his one-person campaign for displaying and using "Student's" table and test of significance in the disputed logic of the *p*-form, will be of interest to more than philosophers and historians of science.

The Insignificance of *p*-Values

Why is *p* so important? Why is *p* the first and last epistemological judge of all statistical observations and experiments? Stolen or not, the *p*-

value does not own a divine or other specially-tailored right to guide scientific judgments at any level of any study and, in fact, at the origins of small sample significance testing, its appointment in science was nothing of the sort.

Closer to the opposite: *p* was almost valueless. Take the early 20th century history of experimental chemistry, agronomy, and brewing, for example. In 1908 a Guinness brewer, the experimental chemist and agronomer William Sealy Gosset (1876-1937), aka "Student," published the original *t*-table (Student 1908, p. 13; ***** 2008). For making probabilistic judgments in the laboratory and field, the inventive "Student," a self-trained statistician, and a business man, did not put much weight on a given test of significance, *t* or *p*: "The latter seems to me to be nearly valueless in itself," he told Egon S. Pearson, then-the editor of *Biometrika* ("Student" 1937, quoted in E. S. Pearson 1939, p. 244). What you really want to know, "Student" told Pearson in 1937, is: (1) the size and nature of your "real error" (not only of your random error), and, (2) Can brewers or CEOs or farmers or whomever "earn a profit" by following the result of the experiment, as if true? Though not the only standards of science, certainly "Student's" idea of statistical knowledge went far beyond simple qualitative consideration of whether $p < .05$—Fisher's dangerous, unprofitable, and needlessly coercive and bureaucratic standard.

Historians haven't much seen these sides of Fisher or "Student," considered in isolation and together. (Egon S. Pearson's notes toward a biography is silent on this matter of the copyright history of *t*: Pearson 1990, posthumous.) Fisher's formidable influence on statistical historiography (Savage 1971) is a forgivable reason but still it is a reason historians don't know much about Fisher's great teacher-by-mail, the anonymous "Student." One hundred years after the *t* table was first presented by William Sealy Gosset aka "Student" (Student 1908), Gosset's pioneering science at Guinness and most of his published articles sit essentially unexamined (Student 1942; Zabell 2008; ***** 2008). This is unfortunate for the history of *p* and indeed for experimental and statistical science in general (List 2008). "Student" was a great scientist who got scooped by another great scientist, Mr. Fisher, one of the most influential scientists of the 20th century.

Stealing *t* took some doing. "Student," as Fisher (1939, p. 9) acknowledged, made with *t* a first-class contribution to the logic and procedures of science. Leaping over Laplace and Lavoisier, between 1905 and 1925, "Student" invented, improved, and extended the *t* table for his own use in situations of small sample estimation and testing (which "Student" also therein invented), copyrighting new and improved versions of his *t* table on four separate occasions: Student 1908, 1914, 1917, 1925. And yet despite "Student's" investments and previous copyrights, Fisher by 1938 would transfer the copyright on *t* to himself. Now "Student's" *t* table gets credited to Ronald A. Fisher (1925) himself, or, nearly as commonly, to Fisher and Frank Yates (1938). After Fisher, "Student's" copyrighted *t* tables (and theories of inference and experimentation) have sat unused, while the credit for "Student's" frequently-reprinted and consulted tables continues to go to Fisher and Fisher and Yates. Is the *p*-form of "Student's" *t* table fraudulent?

On grounds of logic and basic common sense, many reply 'yes, the *p*-form *is* and always *will be* fraudulent' (for example, Zellner 1984, p. 288). The *p*-value is flat-out *unqualified* to be a leading test of science; in truth, the *p*-form of the table subtracts from rather than adds to quantitative understanding in science, because it pretends that *p* is a quantitatively meaningful result. What, after all, *is* p, besides the probability of getting a larger value of "Student's" *t* than the value actually kicked up by your experimental design? As the great Cambridge geophysicist and experimental philosopher Harold Jeffreys (1961, p. 385) observed:

> If P is small, that means that there have been unexpectedly large departures from prediction [from, that is, the assumed-to-be-true null hypothesis of zero treatment or variate effect]. But why should these [departures from the null] be stated in terms of P? The latter gives the probability of departures, measured in a particular way, equal to *or greater than* the observed set, and the contribution from the actual value [of the test statistic] is nearly always negligible. *What the use of P implies, therefore, is that a hypothesis that may be true may be rejected because it has not predicted observable results that have not occurred.* This seems a remarkable procedure. On the face of it the fact

> that such results have not occurred might more reasonably be taken as evidence for the law [the null hypothesis of *no* effect], not against it. The same applies to all the current significance tests based on P integrals.

In other words, the big standard of science is illogical *p*'s. Sadly, few significance testing scientists today follow—or are familiar with—"Student's" and Jeffreys's and others' fundamental epistemological opposition to "significance tests based on P integrals" (***** and ***** 2008, p. 2) and thus to today's use of "Student's" tables and test.

No wonder. Put plainly, today's scientist has a simple incentive to misuse *p*. For example, in "Some Notes on *Science* Style," authors of *Science* are instructed to "Use [the word] 'significance' only when discussing statistical significance." Echoes of Fisher. In the on-line "*Nature* Guide to Authors: Statistical Checklist," it is required that "Actual *P* values are given [in articles published in *Nature*] for primary analyses"—meaning that the most important parts of the investigation are to be reported in the way that Jeffreys and "Student" and basic common sense long ago rejected, that is, Fisher's way.[2] ***** and ***** (2008) find that 80-90% of scientists act in illogical accordance with Fisher's *p*.

Why have so many scientists wandered away from estimation of real error and magnitudes—those "large departures"—to the pseudo-scientific world of calculating *p*-values? The copyright history of "Student's" *t*-table, 1908 to 1938, supplies a part—perhaps a surprisingly large part—of the answer.

“Student” Copyrights His *t*-Table: 1908 to 1925

“Student’s” biochemical-and-econometric problem at Guinness boiled down to a small-sample problem of estimation and prediction for industrial quality control. His job was to statistically control and improve Guinness stout and, with it, the Guinness bottom line (***** 2008). Pounding out up to 100 million gallons of stout per annum, the-then largest brewery in the world needed reliable information.

Faced with persistently small samples, “Student” in 1904 observed that the normal approximation curve was not sufficient for calculating the odds that the results he found had obtained by chance alone. In the Pearson school of statistics, in the early 1900s, large sample sizes and asymptotic results ruled. But experiments are costly to conduct, the beer-brewer noticed, and a good result in the lab or field may not survive on the larger scale of production, requiring additional risky investments of scarce resources. “Student” wished to know how many samples (n) he needed to get odds of, say, 10-to-1 that some result or other obtains (***** 2008).

Once “Student” found a functional form for calculating the probable error of a mean – a feat he accomplished in 1906-1907 during a sabbatical at Karl Pearson’s Biometric Laboratory, in London – the daunting task of tabling *t*’s distribution remained. On “Baby Triumphator,” a mechanical calculator, “Student” computed the first table for $n = 4, 5, \ldots, 10$ (Student 1908a, p. 13). Tabling *t* did not come easily to the-then Head Experimental Brewer of Guinness. The first *t* table took him about six months to compute. (Turning the hand-crank on Baby Triumphator required so much strength that "Student's" assistant, Mr. Somerfield, could not himself power it. "Student," an avid outdoorsman who was married to the captain of the Irish Ladies' National Hockey team, definitely could power it.)[3] Karl Pearson, for Biometrika Trust, was "Student's" editor and publisher for the first three tables of *t*: Student 1908, 1914, and 1917.

The form of "Student's" fourth and last version of *t*, published in Corrado Gini's *Metron*, December, 1925, closely resembles its 1917, 1914, and 1908 precursors (Student 1925): Sample size *n* (with "degrees of freedom") he arranged on the top row—giving "Student" instant access to information about how many experiments he might wish to conduct, depending on the desired odds; *t* values he ran down the left most column; and the theoretical probability of getting a result between $-\infty$ and a given level of estimated *t* he put inside the table itself. But that same year, in Fisher's *Statistical Methods for Research Workers* (1925a), "Student's" table appeared in a *new* form.

Previously hidden contents were plucked out of "Student's" table and given primary emphasis by Fisher. He rearranged "Student's" table – he transposed the axes and inverted the contents of "Student's" table - to emphasize "*the purposes* of the present book" (Fisher 1925a, p. 22), by which he meant the *p*-value and philosophy that would figure so prominently in 20th century science. (*P*- values were to most readers in the 1920s not even discoverable in "Student's" original table.)[4] From "Student's" table Fisher took *n* off the top and ran it down the left most side.

T-values (1.96, 1.64, . . .) he took from the left most side of "Student's" table and put them in the inside rows and columns. *P-values* (.01, .05, . . .) Fisher found by inversion of "Student's" cumulative probability and displayed prominently at the *top* of the table (Fisher 1925a [1928], p. 139 and again in a fold-out Appendix)—as if on the top of science. Fisher called the table "the table of *t*."

On the surface, Fisher's inversion and transposition of "Student's" important table may not seem problematic or even unusual in the history of science. Certainly it's not unusual. When in the 1670s Newton's experimental results on "opticks" failed to persuade Robert Hooke, Christian Huygens, and other critics, Newton - with whom Fisher had more in common than the Royal Society – transformed his experimentally-based arguments into abstract geometric propositions: "Let α represent an oblong piece of white paper," wrote Newton (Newton 1673 quoted in Bazerman 1988, p. 115; cf. Westfall 1980, pp. 244-5). New arguments can demand new forms, and maybe that is all that Fisher, like Newton before him, was doing with the table emphasizing *p* values. He was creating a space for new arguments.

In the early 1890s, the Cambridge economist Alfred Marshall (a teacher of John Maynard Keynes) "transposed the axes" of price and quantity, taking price from the horizontal axis and placing it on the vertical, where it has stayed (S. Gordon 1982). Like Fisher competing to create a new language at the dawn of modern statistics, Marshall had rhetorical reasons for making his new geometric argument. He wished to use supply and demand curves to emphasize the size the overall "welfare" of society—"consumer welfare," Scott Gordon argues, in particular.

Fisher's reasons for transposing the axes appear to be something like the opposite of Marshall's. Fisher, whom some consider to be the most important biologist since Charles Darwin, shied away from making economic and other substantive evaluations and in the mid-1950s he became openly hostile to the economic or "loss function" way of thinking about statistics. Real scientific research is done by an act of "faith," he claimed from positions he held at various elite "counter-laboratories" (Latour 2003, p. 86) to Guinness's Laboratory, run by "Student" himself, over in Dublin (Fisher 1955, p. 75). Speaking of significance tests in conjunction with analysis of variance, regression analysis, null hypothesis test procedures, and the like, Fisher wrote, "We make no attempt to evaluate these consequences, and do not assume that they are capable of evaluation in any currency" (1955, p. 75). No loss functions, no power functions, no economic or epidemiological significance. Just *p*-values, he said, as he quietly erased "Student's" name, work, and arguments from the annals.

From 1925 to 1962 Fisher went to great lengths in lectures and prose work – as shown here and in fuller version in ***** and ***** (2008, chp. 22) – pointing scientists' attention to the top of the new *t*-table, in *p*-form, in various Fisher productions, distributed worldwide.[5] In them Fisher told us to zero-in on the *p* value - "P = .05" (Fisher 1928, p. 139) in particular – and we did.

The evidence suggests that copyrighting the Guinness brewer's table, Gosset's fine table, with the mushy *p's* on top, has conferred more than intellectual rights to R. A. Fisher and Fisher's form of argument.

Fisher's Request, Fisher's Fraud

When Gosset mailed the fourth revision of his *t*-tables to Fisher on September 21, 1922, he exclaimed in the first sentence of a covering letter, "I am sending you a copy of Student's Tables as you are the only man that's ever likely to use them!"[6] Gosset's estimate of who was likely to use "Student's Tables"—and thus "*'Student's' test of significance*" (Fisher 1939, p. 1) — contained a large posterior error. After Fisher, and especially after the arrival of the digital computer, it is difficult to find a scientist who has *not* used Student's Tables.

Gosset's letter replied to a request from Fisher asking for revised and expanded tables.[7] He had no money or co-authorship to offer Gosset in exchange for the new work and tables. Yet Fisher wanted to print the difficult-to-compute tables in the first edition of his *Statistical Methods for Research Workers.*

"Student" had published *t* twice in Karl Pearson's *Biomerika* (in 1908 and, with more *n*, in 1917) and a third time in Pearson's *Tables for Biometricians and Statisticians* (1914). *Tables* was the major reference book before the era of Fisher and Yates's *Statistical Tables for Biological, Agricultural, and Medical Research* (1938), where Gosset's *t* would appear again in Fisher's *p*-form.

Gosset as usual offered to help. He and Fisher would call Gosset's revised tables the "Table of *t*," to acknowledge Fisher's degrees-of-freedom correction to "Student's" 1908 *z,* which was reduced by total sample size.[8] In 1922 bullets and thugs ruled O'Connell Street in Dublin. A strike was on, slowing the Post. Gosset was unusually busy with work at the brewery. Baby Triumphator, his calculator, was needed for overtime work. Fisher would have to wait.

Fisher's tone in reply was urgent. So Gosset put his assistant Somerfield onto the task of preparing the index to Fisher's book. That freed some time for Gosset—who was planning to help with the index—to turn the crank on Baby Trimphator. When he sent the new *t*-values to Fisher, in September 1922, Gosset seemed more concerned with a discussion the two men were having regarding the Poisson distribution (Letter No. 11, Gosset 1962). His only comment about them was the one just noted: "you're the only man that's ever likely to use" the *t* table, Gosset told Fisher. Statistical significance just wasn't very important to Gosset, and perhaps he thought – in 1922 - to Fisher and other scientists. The men discovered some errors in the tables and Gosset went back to work on them.

Fisher then wrote to ask if he (Fisher) could "quote" the completed table. Bells were clanging in Edinburgh in anticipation of Fisher's book-in-preparation at Oliver & Boyd. So Fisher's tone remained understandably urgent. "Dear Fisher," Gosset replied (it is now July 12, 1923), "I think you have all the completed work on the table, but I expect to finish it sometime next winter" (Imagine how the clang of bells must have dinned for just a moment in Fisher's ears). "I should say," Gosset continued, "that it is certainly in course of preparation.

> As to 'quoting' [in Fisher's *Statistical Methods for Research Workers*] the [Gosset] table [previously published] in Biometrika it depends [Gosset told Fisher] just what you

> mean by quoting. I imagine that they have the copyright and would be inclined to enforce it against anyone. The journal doesn't now pay its way though it did before the war and they are bound to make people buy it if they possibly can. I don't think, if I were Editor, that I would allow much more than a reference!
>
> Letter No. 26 in Gosset 1962; emphasis in original.

Fisher eventually got much more than a reference to Gosset's table. But Gosset was not easily letting go of his table (Letter No. 35, 36, 46 in Gosset 1962). Notice, for example, that Gosset underlined, as if to emphasize to Fisher, the "anyone"! The cash-constrained Pearson had been good to Gosset: he published 14 of "Student's" 21 published articles (Student 1942). And Biometrika Trust held the copyright to the table, three-times over. (Gosset was a business man who had donned a pen name in publications, just as Guinness required.) "They [Biometrika Trust] are bound to make people buy" the *t* table, Gosset told Fisher, nudging him to go ahead and do the right thing. But Fisher and his publisher—Oliver & Boyd – could wait. For over a year Gosset worked out bugs in the revised table (Student 1925, p. 107).

Fisher's *p*-Form and Rule

While Gosset was busy fixing beer and *t*, Fisher would – among other things - publish related small sample results in Corrado Gini's journal, *Metron* (in 1921 and 1924). Intended or not, by 1925 Fisher found in Gini an opportunity to muddy the authorship of "Student's" tables. For printing the 1925 tables, *Biometrika* was out, *Metron* was in.

Fisher published two articles in the December, 1925 issue of *Metron*, "Applications of 'Student's' Distribution" and "Expansion of 'Student's' Integral in Powers of n^{-1} "(Fisher 1925b, 1925c). Sandwiched between Fisher's two articles appeared a much shorter article (at three-pages), "New Tables for Testing the Significance of Observations," by "Student." Said he, "[t]he present Tables have . . . at Mr. Fisher's suggestion been constructed with argument $t = z \sqrt{n}$ where n is now one less than the number in the sample, which we may call n'" (Student 1925, p. 106). "Student's" 3-page article explains how "Student" calculated the *t* tables.[9]

What confused readers of *Metron* (and practicing scientists ever after) was the placement of the new tables in *Fisher's* second article (Fisher 1925c, pp. 113-120). According to the physical arrangement of the articles and tables, it appeared to outsiders that Fisher, not "Student," should be credited with *t, p,* and rules about "significance."

Gosset was confused, too. Before the *Metron* publications, and far into 1925, Gosset was still negotiating with Karl Pearson under the impression that Fisher and he wished for a deal with Pearson's *Biometrika.* Gosset was being deceived.

On June 12, 1925, only six months before the completed and printed *Metron,* Gosset explained in a letter to Fisher: "K. P. is very anxious to publish your note [in *Biometrika*] about the use of the table, but doesn't like the binomial approximation which he considers requires a proof of convergence." Gosset told Fisher he remained hopeful that Pearson might publish the tables and three articles.[10] Pearson had "warmed to the proposal."

Fisher "remained keen to retain the right of publishing elsewhere," namely, as he would later reveal, in *Metron.*[11] Correspondence dried up for a while and then from Fisher came this: "I enclose the two notes I mentioned, the first of which is an attempt to give some idea of the multitude of uses to which your table may be put . . . . I have told Oliver & Boyd to send you two proofs [of *Statistical Methods*] as they become available . . . Yours sincerely, R. A. Fisher."[12] Pearson and *Biometrika* and most of all, Gosset, were being quietly skipped over.

Between June 12 and August 10 Fisher secured a deal with Gini and *Metron* (Letter No. 61, 66, Gosset 1962). Or after mid-June he was finally psychologically able to reveal to Gosset his alternative plan. He did and Gosset was mildly annoyed but agreeable with it (all those Pearson copyrights he must have thought). "Dear Fisher, I am sending back your note and my new version which I hope is properly annotated. . . . As to the method of presenting the article [in *Metron*] whether under separate names or joint and the title (Somerfield rather boggles at the title I have put on mine!) I leave it entirely to you to do as you prefer and if necessary to put in liaison Material to putty up the joint."[13] As the statistician and feminist Florence N. David remarked, Gosset possessed "not a jealous bone" in his body (quoted in Reid 1982, pp. 132-3).

Thus empowered Fisher did two things: first, he published the three articles and "Tables" individually, *not* jointly, in the sandwich order Fisher-Student-Fisher-Tables. Second, without doing the "putty up" to make clear the matter of authorship, he published Gosset's tables in his own article, the "Expansion" article (Fisher 1925c, pp.

113-120). In Gini's journal Fisher arranged for the collaboration a Gini coefficient, so to speak, of extreme inequality between him and Gosset.

The "Student" name faded fast from that point forward. Again no surprise. Consider Fisher and Yates's *Statistical Tables for Biological, Agricultural, and Medical Research* (1938). Fisher and the as yet un-reconstituted Yates failed to thank "Student" and *Biometrika* for permission to reprint "The Table of *t*" (v-viii, p. 46; cf. Yates 1951). "Student's" four copyrighted tables of *t* are not referenced in any of Fisher and Yates's 82 references. Less than a year after "Student's" death, "Student's" priority and tables went missing, and with them his actual approach to estimation and hypothesis testing. Claimed Fisher and Yates (1938, p. 1):

> Tables I to V and VII constitute a group of tables, based on the normal distribution, and now widely used in making tests of significance ["The Table of *t*" is Table III]. The common uses of these tables are fully illustrated with numerical examples in Fisher's *Statistical Methods for Research Workers*, where they were first published.

Poor, neglected "Student."

Fisher created a fictional "Student," an experimentalist who was allegedly entranced by *p*-values, randomization, and fixed rules of significance. The actual "Student" was nothing like Fisher's significance tester (***** 2008). "Student" computed odds ratios to assist in what always remained to "Student" an *economic* question of how much effect (1908, p. 22). Said "Student" to Karl Pearson, as early as 1905:

> When I first reported on the subject [of "The Application of the 'Law of Error' to the Work of the Brewery" (1904)], I thought [he told Karl Pearson] that perhaps there might be some degree of probability which is conventionally treated as sufficient in such work as ours and I advised that some outside authority in mathematics [such as Pearson] should be consulted as to what certainty is required to aim at in large scale work. However it would appear that in such work as ours *the degree of certainty to be aimed at* must depend on the *pecuniary advantage to be gained by following the result of the experiment, compared with the increased cost of the new method, if any, and the cost of each experiment*.[14]

Ignoring Gosset's economic approach to the logic of uncertainty, it was Fisher who told and re-told generations of scientists how to treat "significance" and *p* in judgments of experimental output:

> The value for which P=.05, or 1 in 20, is 1.96 or nearly 2; *it is convenient to take this point as a limit in judging* whether a deviation is to be considered significant or not. Deviations exceeding twice the standard deviation are thus *formally regarded* as significant (Fisher 1925a [1941], p. 42, italics supplied).

> Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and *ignore entirely all results which fail to reach this level* (Fisher 1926b, p. 504)

> It is principally by the aid of [significance testing] that these studies may be raised to the rank of sciences (Fisher 1925a [1928], p. 2).

> Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis (Fisher 1935, p. 16).

> In fact, scientific research is not geared to maximize the profits of any particular organization, but is rather an attempt to improve public knowledge undertaken as an act of faith . . . *We make no attempt to evaluate these consequences, and do not assume that they are capable of evaluation in any currency* (Fisher 1955, p. 75, italics supplied)

If Fisher ever confessed to the shift in perspective caused by the *p*-form and rule it was in a single oblique sentence printed in early editions of *Statistical Methods*: “For the purposes of the present book we require the values of *t* corresponding to *given* values of *P* and *n*” (Fisher 1925a [1928, 2nd edition], p. 22; emphasis added). The purposes-of-*P* sentence was deleted from later editions and reprints of Fisher’s book, leaving “Student” and “Student’s” tables out in the scientific yard chirping with

Jeffreys, Neyman, and other "forgettable" birds (Fisher 1955, 1956). How much profit Oliver & Boyd earned from sales of "Student's" *t* is anyone's guess. The intellectual capital amassed to Fisher is something large (and largely observable).

Meantime, Fisher's fraud unknowingly endures in textbook after textbook, journal after journal. Even Bayesian re-printers of "Student's" table express thanks for sacrifices made by "the Literary Executor of the late [anti-Bayesian] Sir Ronald A. Fisher, F.R.S., [and] to Dr. Frank Yates, F.R.S." (DeGroot 1975 [1989], v). In 2005, yet another distinguished statistician thanked Fisher and Yates for permission to reprint "Table III of Fisher and Yates ["Student's" *t* table]. . . published by Oliver and Boyd, Edinburgh, and by permission of the authors and publishers" (S. James Press 2005, pp. ix, 533). "Student," meantime, all but vanished.

Acquiring copyrights to "Student's" table was probably not necessary for another scientist to claim *intellectual* rights to "Student's" test. But it may have been sufficient.

Great scientists, surprise, surprise, are not virtuous in every regard. Pasteur advanced a theory of rabies (and famously treated a young French boy) despite Pasteur's lack of experience with the treatment, the disease, and the boy (Geisson 1995). Lavoisier, a founding father of quantitative chemistry, couldn't compute the size and distribution of his experimental and random errors. ("Student" appeared later, and showed us how we could.) And Lavoisier and Pasteur both misled the public—more than once—about the frequency and nature of their experimental trials (Holmes 1985). Fisher, though a leading man of science, wasn't perfect, either, as

MacKenzie (1981) and others have observed. Fisher was an ingenious theorist who took imaginary long-run frequencies and random ethical rules into his own hands. With his fables about "Student's" *t* he took ambition in science to a different level—the *p* level.

REFERENCES

Altman, Douglas G. 1991. "Statistics in Medical Journals: Developments in the 1980s." *Statistics in Medicine* 10: 1897-1913.

Bazerman, Charles. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science.* Madison: University of Wisconsin Press.

Box, Joan Fisher. 1978. *R. A. Fisher: The Life of a Scientist.* New York: Wiley.

Burke, Kenneth. 1950. *A Rhetoric of Motives.* Berkeley: University of California Press.

DeGroot, Morris H. 1989. *Probability and Statistics*. Reading, MA: Addison-Wesley

Fisher, R. A. 1925a [1928]. *Statistical Methods for Research Workers*. Second edition. Originally published at Edinburgh, Oliver & Boyd, in many reprints, editions, and languages.

Fisher, R. A. 1925b. "Applications of 'Student's' distribution." *Metron* V(3, Dec.): 90-104.

Fisher, R. A. 1925c. "Expansion of 'Student's' Integral in Powers of n^{-1}." *Metron* V(3, Dec.): 110-120.

Fisher, R. A. 1926. "Arrangement of Field Experiments." *Journal of Ministry of Agriculture* XXXIII: 503-13.

Fisher, R. A. 1935. *The Design of Experiments.* Edinburgh: Oliver & Boyd. Reprinted in eight editions and at least four languages.

Fisher, R. A. 1939. "'Student'." *Annals of Eugenics* 9: 1-9.

Fisher, R. A. 1955. "Statistical Methods and Scientific Induction." *Journal of the Royal Statistical Society, Series B (Methodological),* Vol. 17, No. 1, pp. 69-78.

Fisher, R. A. 1956 [1959]. *Statistical Methods and Scientific Inference.* New York: Hafner. Second edition.

Fisher, R. A., and Frank Yates. 1938 [1963]. *Statistical Tables for Biological, Agricultural and Medical Research.* Edinburgh: Oliver and Boyd. Sixth edition.

Freiman, Jennie A., T. Chalmers, H. Smith, R. R. Kuebler. 1978. "The Importance of Beta, the Type II Error and Sample Design in the Design and Interpretation of the Randomized Control Trial: Survey of 71 "Negative" Trials." *New England Journal of Medicine* 299: 690-4.

Geisson, Gerald. 1995. *The Private Science of Louis Pasteur*. Princeton: Princeton University Press.

Gigerenzer, Gerd, Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty, and Lorenz Kruger. 1989. *The Empire of Chance*. Cambridge: Cambridge University Press.

Goodman, S. 1999a. "Toward Evidence-based Medical Statistics. 1: The *p*-value Fallacy. *Annals of Internal Medicine* 130: 995-1004.

Gordon, Scott H. 1982. "Why did Marshall Transpose the Axes?" *Eastern Economic Journal* 8(1, Jan.): 31-45.

Gosset, William S. [see Student]. 1936. "Co-Operation in Large-Scale Experiments." *Supplement to the Journal of the Royal Statistical Society* 3(2): 115-36.

Gosset, William S. 1962. *Letters of William Sealy Gosset to R. A. Fisher. Vols. 1-5.* Eckhart Library, University of Chicago. Private circulation (about 70 copies worldwide).

Holmes, F.L. 1985. *Lavoisier and the Chemistry of Life*. Madison: University of Wisconsin Press.

Jeffreys, Harold. 1961. *Theory of Probability*. Third revised edition. London: Oxford University Press.

Lang, Janet M., Kenneth J. Rothman, and Cristina I. Cann. 1998. "That Confounded *P*-Value," *Epidemiology* 9 (1, Jan.): 7-8.

Latour, Bruno. 2003 [1987]. *Science in Action.* Cambridge: Harvard University Press.

Lehmann, Erich. 1999. "Student" and Small-Sample Theory," *Statistical Science* 14 (4): 418-26.

List, John A. 2008. "Homo experimentalis evolves." *Science* 321 (no. 5886): 207-8.

MacKenzie, Donald A. 1981. *Statistics in Britain, 1865-1930: The Social Construction of Scientific Knowledge.* Edinburgh: Edinburgh University Press.

Pearson, Egon S. 1939. "'Student' as Statistician." *Biometrika* 30 (3/4, Jan.): 210-50.

Pearson, Egon S. 1990 [posthumous]. *'Student': A Statistical Biography of William Sealy Gosset.* Oxford: Clarendon Press. Edited and augmented by R. L. Plackett, with the assistance of G. A. Barnard.

Pearson, Karl. 1914. *Tables for Statisticians and Biometricians, Part I* (London: Biometrics Laboratory, University College London).

Press, S. James. 2003. *Subjective and Objective Bayesian Statistics.* New York: Wiley.

Reid, Constance. 1982 [1998]. *Neyman: A Life.* New York: Springer.

Reiss, Julian. 2008. *Error in Economics: Towards a More Evidence-based Methodology*. London: Routledge.

Savage, Leonard J. 1971a [1976 posthumous]. "On Re-Reading R. A. Fisher." *Annals of Statistics* 4(3): 441-500.

Student. 1908a. "The Probable Error of a Mean." *Biometrika* VI (1, March): pp. 1-24.

Student 1914. (See Pearson, Karl: 1914.)

Student. 1917. "Tables for Estimating the Probability that the Mean of a Unique Sample of Observations Lies Between $-\infty$ and Any Given Distance of the Mean of the Population from which the Sample is Drawn." *Biometrika* 11(4, May): 414-17.

Student. 1925. "New Tables for Testing the Significance of Observations." *Metron* V(3, Dec.): 105-108.

Student. 1942 [posthumous]. Student's *Collected Papers* (London: Biometrika Office). Eds. E. S. Pearson and John Wishart.

Westfall, Richard. 1980 [1995]. *Never at Rest: A Biography of Isaac Newton.* Cambridge: Cambridge University Press.

Yates, Frank. 1951. "The Influence of *Statistical Methods for Research Workers* on the Development of the Science of Statistics," *Journal of the American Statistical Association* 46**:** 19-34.

Zabell, Sandy. 2008. "On Student's 1908 Article," *Journal of the American Statistical Association* 2008 (103): 1-7.

Zellner, Arnold. 1984. *Basic Issues in Econometrics*. Chicago: University of Chicago Press.

***** . 2008. "Guinnessometrics: The Economic Foundation of 'Student's' *t*," *Journal of Economic Perspectives*, Vol. 22, forthcoming Fall 2008.

***** and *****. 2008. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives* (Ann Arbor: University of Michigan Press, 2008).

Biography

********** is currently *********************. He has published numerous articles on the history and philosophy of statistical significance testing in economics and other sciences. His most recent book is **.

[1] Evidence on how *p* gets used and interpreted in the history, philosophy, and practice of science is reported and discussed by ***** and ***** 2008.
[2] In *Science*: http://www.sciencemag.org/about/authors/prep/res/style.dtl; in *Nature*: **http://www.nature.com/nature/authors/gta/index.html**
[3] Letter of Gosset to Fisher, reprinted in Box 1978, p. 117.
[4] See, for example, Gigerenzer, et al., 1989, *The Empire of Chance.*
[5] Fisher, p. 42, *passim*; Fisher 1925a, p. 121; See also ***** and ***** 2007, chps. 21-22.
[6] Letter No. 11 in Gosset 1962; cf. E. S. Pearson 1990, p. 49.
[7] *Ibid*, pp. 49-53.
[8] Student 1925, p. 106; Fisher 1925a, p. 174; Fisher and Yates 1938, p. 46.
[9] Another Gosset fact: In two letters of May, 1926 Gosset gave the ideas of power and decision to Egon S. Pearson. See for example Letter No. 1 of W. S. Gosset to E. S. Pearson, May 11, 1926, in Pearson Papers, Egon file, Green Box, UCL. Egon Pearson always credited "Student" for the power-idea: for example, E. S. Pearson 1939, p. 243.
[10] Letter No. 47, Gosset 1962.
[11] Pearson 1990, p. 51.
[12] Letter No. 48, in Gosset 1962.
[13] *Ibid*, Letter No. 66.
[14] Gosset, c. April 1905, in E. S. Pearson 1939, pp. 215-216.