



David Aronson: STRUCK BY SCIENCE

Despite years of success as a technician, losses convinced this trader and author that technical analysis is full of holes. His solution combines objective testing with statistics.

BY DAVID BUKEY

David Aronson had studied technical analysis for 40 years and was a true believer. However, his faith in the effectiveness of technical analysis was shaken in the summer of 2000, just after the technology bubble burst.

At the time, Aronson, 61, worked as a proprietary trader for Spear, Leeds, and Kellogg and had been profitable for four years. He used standard technical analysis techniques — interpreting classic chart patterns and standard indicators such as the Relative Strength Index (RSI), which he now describes as “seat-of-the-pants analysis.”

After losing money in the first half of 2000, Aronson compared his performance since 1996 to the Nasdaq 100 index and found surprising results.

“My alpha — the difference between my performance and the benchmark — was basically zero,” he says. “I was floating



up with the market.”

Although his managers weren't bothered by this fact, Aronson concluded that subjective approaches to technical analysis, such as chart patterns, Gann, and Elliott Wave analy-

sis, were deeply flawed.

These methods, Aronson realized, weren't testable, which means they couldn't be quantified and evaluated objectively. For instance, a technician may find a head-and-shoulders pattern on a price chart, but the identification process is completely subjective. Such patterns lack precise rules that define them objectively and can be used to determine whether they have true predictive power.

"It's like a bogus health claim: If you wear a copper bracelet you'll feel better and your golf game will improve," Aronson argues. "Well, how will you feel better? How much will your game improve? They don't make testable assertions."

The answer, according to Aronson, was to apply the scientific method to technical analysis and change it from a pseudo-

The final section of *Evidence-Based Technical Analysis* describes Aronson's test of 6,400 trading rules on the S&P 500 index since 1980. He then applies various statistical tests to determine whether the best-performing rules met his strict guidelines. The test results show how difficult it is to find truly significant strategies — an invaluable lesson.

Over the years, Aronson has worked as a stock broker and co-founded two businesses: AvoCom Corp., formed in 1979 to analyze performance data on commodity fund managers, and Radon Research Group, which he started in 1982 to develop adaptive pattern-recognition software.

These days, Aronson teaches a class on technical analysis and data mining at Baruch College in New York and continues to trade and analyze various markets.

It never occurred to me that those early successes may have been luck. I didn't start to realize the role of randomness in markets until much later.

science such as astrology into a true science such as astronomy.

However, simply testing objective rules against historical price data (back-testing) doesn't prove a rule will be profitable in the future. You must also use statistics such as significance tests and confidence intervals to prove past performance isn't just luck.

Combining statistics and the scientific method intrigued Aronson. Although he never formally studied statistics, he hired a tutor and dove into the subject in the summer of 2001.

"It was a revelation," he says. "I saw the connection to the scientific method and thought, this is way technical analysis needs to be approached."

The result is Aronson's new book, *Evidence-Based Technical Analysis* (John Wiley & Sons, 2006), which critiques subjective approaches to technical analysis, describes the biases that make their claims seem valid, and explains why objective testing combined with statistics are logical answers to the challenges facing traders.

His book also examines the pitfalls behind back-testing multiple trading rules and selecting the best-performing one ("data mining"). A strategy's past performance always contains an element of luck, but testing many rules and picking the best one increases the odds that "superior" performance is the result of luck. To guard against this, Aronson describes how to incorporate two approaches — bootstrapping and Monte Carlo analysis — that take data-mining bias into account and gauge the statistical validity of trade results.

AT: *When did you become interested in the financial markets?*

DA: I learned about stock investing when I was 13. When I was 14 or 15, I read Nicholas Darvas' *How I Made \$2 Million in the Stock Market*. That was my first exposure to technical analysis. Plotting the moves and making trades based on price really fascinated me. Darvas' box pattern was a common technical idea — support and resistance levels, rectangular formations, and so on.

The point-and-figure method was my next discovery. My brother and I went to the library, studied the papers, and created our own point-and-figure charts. In high school, I gave stock tips to teachers. My first couple of chart picks went well. I found a triple-top buy signal in Cubic Corp., and the stock climbed 10 points within three or four days. I looked like a magician.

However, that didn't last too long — my track record faded. But it never occurred to me those early successes may have been luck. I didn't start to realize the role of randomness in markets until much later.

AT: *I noticed you have a philosophy degree. Did your interest in the markets fade while you were in college?*

DA: No. I even met with the investment club at Lafayette College and tried to convince them that technical analysis was

continued on p. 46

Evidence-Based Technical Analysis:

Applying the Scientific Method and Statistical Inference to Trading Signals

By David Aronson
John Wiley & Sons, 2006
Hardcover, 528 pages
\$95

David Aronson's *Evidence-Based Technical Analysis* questions the validity of subjective approaches to technical analysis. Instead of studying the unverifiable claims of W.D. Gann and Ralph Nelson Elliott, among others, Aronson argues technicians should focus on objective price patterns and strategies that can be tested. However, simply testing rules against historical price data isn't enough; these results must then be measured with statistics.

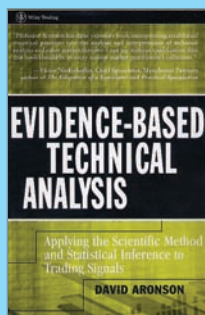
Evidence-Based Technical Analysis isn't just an indictment of subjective techniques. Aronson explains why so many technical methods are difficult to verify, and he describes several biological biases that contribute to the illusion that classic chart patterns such as head-and-shoulders, pennants, or triangles have predictive power.

He then explains in detail how traders should analyze the markets by combining the logical framework of the scientific method with statistical tests. Finally, Aronson presents a case study of 6,400 technical rules or strategies on the S&P 500 over the past 25 years to illustrate his statistical approach.

One of the book's targets is "data mining," which is the process of testing multiple rules and selecting the top-performing one. While Aronson encourages traders to use this method to find profitable rules, he also warns against data-mining bias (the more rules you test, the greater the role of luck). To combat this problem, Aronson introduces two solutions — bootstrapping and Monte Carlo analysis — which measure each rule's performance and determine whether the "best" rule's performance is statistically significant.

The argument is well-researched and lengthy, but never dull. Aronson discusses the philosophical foundations of science with insight and introduces statistical concepts with helpful tables and charts — one of the most straightforward explanations of this subject you'll find.

Although Aronson describes his trading experiences in only one section of the book, his story adds a compelling element. He relied on subjective approaches to technical analysis for decades before realizing their flaws. At first glance, *Evidence-Based Technical Analysis* may resemble a textbook, but it is really a story of one trader's honest attempt to find valid and profitable price patterns.



the head of technical analysis at the firm. On the basis of cycle analysis, the market was about halfway through the bear market that started in January 1973. I predicted the market would hit bottom sometime in the fall of 1974, which occurred as predicted. Again, maybe it was luck.

AT: *Is that how you view that market call and your past successes with technical analysis in the context of your book's theme — that subjective approaches aren't testable and can be completely off-base?*

DA: While I was very excited about technical analysis, [I also had some doubts]. My friend John Wolberg, a nuclear engineer, had access to large computers. So one night in 1976, we entered all monthly Dow Jones Industrial Average prices going back to 1898. We used the same cycle-analysis techniques that Hurst used and found the cycles really weren't there. John got a big kick out of the fact that Hurst's cycle theory was [simply an alternate] description of the market's long-term uptrend. However, I don't think it quite registered with me.

When I was at Merrill Lynch, I also learned about trading commodities with objective trend-following methods. I was immediately attracted to back-testing rules because they gave definitive signals.

I left Merrill in 1979 to work at Douglas Stuart, a small brokerage firm. I

studied the track records of CTFC-registered trading advisors, and I learned many were using similar trend-following systems. But a couple of firms were doing something remarkably different — using artificial intelligence, or data mining, to develop profitable strategies. They weren't just applying some human-devised rule, they were allowing computers to *discover and create* rules.

superior to fundamental analysis. I was laughed out of the club.
After graduating in 1967, I went to law school for a year. The Vietnam War was going on, so I joined the Peace Corps. I went to El Salvador for a couple of years and came back in 1970. Although I wanted to work for a brokerage firm, stocks had just finished a two-year bear market, and I was unable to get a job on Wall Street.

When the market came back in 1973, I took a job at Merrill Lynch as a stock broker. At the time, Jim Hurst's book *The Profit Magic of Stock Transaction Timing* was making a big splash. I got some other brokers interested in Hurst's cycle analysis — it seemed very scientific.

After I finished Merrill Lynch's training school in December 1973, I studied cycle analysis and wrote a memo to Bob Farrell,

studied the track records of CTFC-registered trading advisors, and I learned many were using similar trend-following systems. But a couple of firms were doing something remarkably different — using artificial intelligence, or data mining, to develop profitable strategies. They weren't just applying some human-devised rule, they were allowing computers to *discover and create* rules.

AT: *Is this similar to a voting system for multiple rules?*

DA: There are multiple rules involved, but it is not a simple voting system. In the voting system, let's say you have 10 different rules. If seven rules are positive and three are negative, that's a buy signal. That method of integrating different signals is a linear method — and there's nothing wrong with that.

These more advanced machine-learning methods gave

many different indicators to the computer and let it combine them. This seemed to have a greater potential than simplistic trend-following methods.

For example, I read an article in an engineering journal that discussed a new method of cooling steel when it comes out of a blast furnace. When you make hot, rolled steel, you have to cool it at a very precise rate. Typically, workers did this manually — they'd guide cooling sprays onto the steel to cool it properly. If you cool it too fast or too slow, you will ruin it.

This article advocated a new approach: take measurements in the blast furnace and [build a] machine to predict temperature and control the cooling spray. This technique was another example of taking information from many variables and combining it into a prediction. The markets don't behave as well as the reaction inside of a blast furnace, but it's a similar problem.

In the early 80s, I co-founded two businesses: AdvoCom Corp., which built portfolios of successful commodity trading advisors for investors, and Radon Research Group, which developed a suite of pattern-recognition software tools and helped traders develop prediction models. Radon's main business was improving the performance of existing trading systems, so we filtered their signals.

AT: *Did you evaluate the performance of many rules to find the best one?*

DA: Let's say we were trying to improve a trading system's signals. We didn't need to know the system, we just needed a long history of its performance.

Rather than building a model from scratch, we'll try to find a model that has some positive performance and improve it. This was a novel application of data-mining techniques to technical analysis. We consulted for Paul Tudor Jones, Manufacturers Hanover Bank, and a Bermuda-based oil company that traded actively.

AT: *Were your techniques successful?*

Statistical significance and confidence intervals

When traders evaluate historical trade performance for **statistical significance**, they're trying to determine whether or not past performance differs enough from the hypothesis that the results have a future expected return of zero (known as the "null hypothesis").

The test determines the probability that the back-test results are truly different from a flat return. For example, if a certain trading rule produced a 12-percent profit, is that return large enough to reject the null hypothesis that the rule has no predictive value? If the rule's performance has a low probability — 1 to 5 percent — of occurring from luck alone, the performance is considered "statistically significant."

Although there are several ways to test for significance, the T-test is a standard formula:

$$T = \text{sample mean} - \text{population mean} / \text{standard error of the mean}$$

where,

Sample mean = the rule's average performance

Population mean = the future expected performance (0)

Standard error of the mean = standard deviation / sqrt(sample size)

Finally, look up T's value and the sample's degrees of freedom (sample size - 1) in a T-table (see www.statsoft.com/textbook/sttable.html). If T is higher than the table's value, the performance was significant enough to reject the null hypothesis.

Confidence intervals offer a more precise estimate of a trade rule's expected performance by setting upper and lower limits for its central tendency. This range can suggest how precise a statistic (mean, correlation, etc.) is relative to a certain probability. For example, 98-percent confidence levels suggest a certain result will likely fall outside of the designated range just 2 percent of the time.

Assume a trading rule's mean (average) return is 17.58 percent, and you want to find the upper and lower confidence levels for this mean at the 95-percent confidence interval. Let's say there are 12 individual returns of 12, 15, 17, 23, 45, 33, 67, -24, 11, 10, 21, and -19 percent.

The standard deviation of these returns is 24.64 percent, and you must go 1.96 standard deviations above and below the mean to capture 95 percent of the returns. (The following formula uses a "z value" of 1.96, representing the x-axis distance of either side of the mean.)

The standard error of the mean is 7.11 percent (24.64 percent / sqrt(12)). The formula for the 95-percent upper and lower confidence levels are:

Lower limit = mean - (1.96 * 7.11 percent standard error) =
17.58 percent - (1.96)(7.11 percent) = 3.64 percent.

Upper limit = mean + (1.96 * 7.11 percent standard error) =
17.58 percent + (1.96)(7.11 percent) = 31.5 percent.

This means there is a 95-percent likelihood the expected returns will be between 3.64 and 31.5 percent. (For more information on standard deviation, see "Key concepts" on p. 77.)

DA: Well, we didn't really know. After we delivered the filter, we didn't get to see the results. However, once we developed a filter for a client with a basic breakout system for T-bonds and tracked the results for about a year. He was quite successful.

continued on p. 48

AT: *Did you work at Radon until 1996 when you became a proprietary trader at Spear, Leeds, and Kellogg?*

DA: Yes. Radon Research Group continued until 1999, after I became a prop trader. When we started Radon in 1982, our software was fairly unique. However, by the mid-90s, there were many neutral-network packages available, and we couldn't compete.

AT: *What was your experience as a proprietary trader?*

DA: A friend of mine worked at Spear, Leeds, and Kellogg and asked me to join them. I didn't have any systems, but he suggested that I just analyze charts. In 1996 I was a chartered market technician, but I hadn't realized that subjective methods of technical analysis lacked validity. I liked testing objective rules with computers, but I decided to analyze the charts and do the best I could.

I also came across two books that really resonated with me: Thomas Gilovich's *How We Know What Isn't So* and Michael Shermer's *Why People Believe Weird Things*.

AT: *Is that when you realized the connection between the scientific method and subjective technical analysis?*

DA: Right. During that time, I started to develop an objective system. The basic idea was to create a filter for a trend-following system in stocks. I hired a programmer, but I had limited time and resources.

Then, I asked the firm to buy special back-testing software, which included preprogrammed systems that showed great results. But how did this program find highly profitable systems?

The software had many different trading models, which it applied to thousands of stocks. For instance, it advertised a

People who believe in subjective technical analysis don't realize it's an empty claim. Once you entertain the possibility that an untestable claim might be true, it's very easy to find seeming confirmation. We can all find head-and-shoulder patterns and say "Wow, it worked here."

AT: *What aspects of technical analysis did you focus on?*

DA: Classic chart patterns and subjective evaluation of various technical indicators.

AT: *Were you back-testing ideas at this point?*

DA: No. However, every time I entered a trade, I'd note the technical rationale for it. I also had a predetermined stop-loss. After exiting a trade, I would really analyze it. As the market rose from 1996 to 1998, I had a bullish bias and was profitable.

However, in the summer of 2000, I compared my monthly performance to the Nasdaq 100 and found they were nearly identical. My colleagues weren't particularly quantitative, so I don't think my problem registered. My manager said "You're making money. That's what we're interested in." However, I felt I had a duty to let them know I was not adding value.

When the market turned down and I didn't adapt my trading style, I gave back most of my gains between the spring of 2000 and the spring of 2001. I started thinking about writing this book, because my own subjective trading had not worked.

AT: *Did keeping a trade journal help your trading?*

DA: It helped destroy my faith in subjective technical analysis. I chose trades carefully, and I seemed to be randomly correct.

model that worked for JDS Uniphase. However, they picked that result out of thousands of stocks, and they held out JDS Uniphase's [returns] as indicative. I didn't really have an in-depth understanding of the data-mining bias then (*the more rules you test, the greater the odds of luck entering the picture*), but I knew intuitively that JDS Uniphase's cherry-picked performance didn't make sense.

It's simply too easy to believe things that aren't true, particularly in highly random, complex financial markets. I needed to dig into formal statistics, which I never really studied in school. The whole time I ran Radon Research Group, I never mastered the basic foundations of statistics. I liked the idea of machines learning, and they were conducting statistical tests internally, but I didn't completely understand some of those key concepts.

AT: *You also discuss human biases that help hide the truth behind subjective technical approaches. Do you mind explaining some of the biases that interfere with finding sound trading methods?*

DA: First, people who believe in subjective technical analysis don't realize it is an empty claim. Once you start entertaining the possibility that an untestable claim might be true, it's very easy to find seemingly confirmatory evidence. We can all find

anecdotes of head-and-shoulder patterns and say “Wow, it worked here.”

AT: *Is the hindsight bias related to this? For example, looking at historical charts and saying, “Of course the pattern was bearish because the market dropped afterward.”*

DA: Yes. Again, when you examine charts for subjective patterns, you can easily find them after the fact. When you look at historical charts, the market’s post-pattern behavior is staring at you. This makes a market seem more predictable than it really is.

Several types of biases help confer the illusion of validity on subjective methods, which are untestable claims. The only way to avoid such erroneous beliefs is to use the scientific method. The scientific method was used for hundreds of years before it was realized exactly why it worked as well as it did. Philosophers such as David Hume and Karl Popper helped explain why it is so effective. One of Popper’s realizations was that you cannot use evidence to prove a hypothesis true.

For example, the following statement is true — “If it’s a dog, then it has four legs.” If I then see a creature that has four legs, it is fallacious to conclude it must be a dog. Similarly, it is wrong to argue a [strategy] is valid simply because we found profitable examples. That commits the same mistake. The method might be good, but not necessarily.

Instead, the scientific method is based on indirect proof. Evidence can be used to falsify a hypothesis. So if we start with the hypothesis that a tested method does not work (null hypothesis), it is possible to use evidence of profits to falsify it. This indirectly proves the method generates profits (alternative hypothesis). This is the logic on which the scientific method rests.

AT: *It seems serious traders would agree that subjective methods are untestable. But what’s wrong with simply testing an objective rule or strategy? Why can’t I accept a strategy’s back-tested performance?*

DA: First, how did you find that rule? Did you look at many rules, or was it the first rule tested? In other words, were you data mining or just testing a single hypothesis?

Let’s say you just tested one rule and it was profitable. It’s still possible that performance was just luck. You need to exclude this possibility before concluding the rule really has merit.

For instance, when Jonas Salk tested his polio vaccine on a sample of patients, he really wanted to know how well the vaccine would work in the general population. Salk needed statistically significant evidence — an infection rate that was so much lower in his sample that it was unlikely to have occurred by chance.

People taking the real vaccine did have a much lower rate of polio infection than the placebo group. That difference was large enough that it was unlikely to have happened by chance.

A trader faces the same situation. He may have just found a rule that was lucky in the historical sample. It is more reasonable to expect luck than to find the Holy Grail. So, you must test observed performance for statistical significance.

Let’s say the back-tested rule gains 50 percent per year. Statistics can tell us the likelihood of a truly worthless rule earning that much in a back-test.

Let’s say the sampling distribution says there is only a 1-in-10,000 chance that 50-percent return was lucky. Then you can reject the null hypothesis and accept the alternative hypothesis — the rule has an expected return of greater than zero.

Many strategies that traders are using would not hold up if examined correctly.

AT: *Do you need to test for significance and find the confidence levels — the likely range of the strategy’s performance — before accepting it as valid?*

DA: They are two different things. The significance test answers a yes-no question: Was it lucky or not? The confidence interval measures the rule’s likely return.

AT: *Most back-testing software packages I’ve seen don’t test results for statistical significance or confidence intervals. How can traders use these concepts?*

DA: If they’re just testing a single rule, then any conventional statistical package with a T-test or confidence interval will work. For example, in my class, we just tested a rule invented by Dr. Martin Zweig. Using the assumption Zweig didn’t engage in data mining, there was a high probability the rule’s performance wasn’t due to luck. My students used a formula straight from a statistics book.

However, these days, most researchers test and tweak rules repeatedly. In this case, a conventional T-test will not work. There are two [additional] tests I suggest in the book — the Monte Carlo permutation method and White’s Reality Check. Those methods take the back-tested results of every rule and use a different technique to construct the best-rule’s probability distribution.

AT: *Do you mind explaining the data-mining bias and how traders can avoid it?*

DA: Well, I don’t think traders should avoid data mining. I encourage them to consider many rules and select the best one. But when you test many rules and say, “Out of 700 rules, the

continued on p. 50

28th one tested performed the best, so I will trade it," you may be making a big mistake. It is very likely that performance has a big component of luck.

Take a different problem where randomness is not a big factor: Finding a musician for an orchestra. Here, the observed performance is mostly merit. Maybe the musician is having a bad day. However, luck isn't really a factor when someone who has practiced for tens of thousands of hours with an instrument. Skill and training drives the performance. When you compare 700 musicians, it is highly likely the best performing musician is excellent and better than everybody else.

Testing a rule is different. Ideally, a rule has some predictive power, but when we pick a top rule among many, its positive

cal significance?

DA: It's a beautiful example of the over-confidence bias, because I thought at least one of the rules would have some merit, but none did. If I used a conventional test, which is appropriate for testing one rule, it would have been hugely significant. The book's test illustrates the importance of these advanced tests when you're data mining.

AT: *Are you worried readers might find the test results anticlimactic — that the book describes these statistical methods and the case study didn't find significant patterns? Or is the testing process itself as valuable as finding a solid rule?*

DA: If someone's buying this book to find the golden fleece,

Technical analysis will fade unless it becomes a real science that ties itself to objective, testable methods and scrutinizes results.

performance includes a great deal of luck. Let's say you have 30 rules that each have an expected return of zero. But on any sample of data, performance will vary randomly above or below zero. If one rule managed to gain 60 percent, data miners are going to pick that one, [and be misled].

AT: *The more rules you consider, the larger the data mining bias?*

DA: Right. If you look up enough lottery players, you'll find someone who has won twice in a row.

AT: *I assume there are more complex statistical tests to find out whether the hand-picked rule's performance is significant?*

DA: To explain how this works, I tested 400 artificial rules that I knew had no predictive power. Because each rule had a slightly different sample of data, its performance wasn't zero; the best rule performed better than that.

When I repeated this experiment 10,000 times, I found the best rule of 400 useless ones gained 48 percent annually, on average. This test was based on a 24-month performance history of the S&P 500 going back to 1928.

AT: *In order for that return to be statistically significant, it might have to be 60 or 70 percent?*

DA: Right.

AT: *Could you interpret the book's case study results for me — testing 6,400 technical analysis rules on the S&P 500, selecting the best performing one, and testing it for statisti-*

cal significance? But the more fundamental question is, "How do you distinguish real gold from fake gold"?

It would have been nice to find some gold nuggets. But my case study proves how difficult it is to find significant rules. Many strategies traders are using would not hold up, if examined correctly.

AT: *So it's wrong to conclude that technical analysis can't find significant rules?*

DA: Yes. In the book, I discuss other studies that show technical analysis has value. A study of 36,000 rules found strategies that work, and trend-following methods in futures markets have been successful, even though they're having a rough time lately. It's just very difficult.

Many books on technical analysis mislead readers by making it look easy. They list rules that supposedly work, but they don't offer evidence.

AT: *At the end of Evidence-Based Technical Analysis, you state technical analysis will be marginalized to the extent it doesn't modernize. What do you mean?*

DA: For example, astrology began as the non-scientific study of the heavens. But the science of astronomy emerged. Now, astrology has been marginalized by scientific practice.

Technical analysis was developed by people with a non-scientific viewpoint — myself included. It will fade unless it becomes a real science that ties itself to objective testable methods and scrutinizes results. It's already happened to a certain extent. We are at a crossroads and technicians can decide to go this new route and turn it into a rigorous science or not. 📌