The Numbers of our Nature

IS THERE A MATH OF STYLE?

BY DAN ROCKMORE

In May 2005, Alex Matter, a filmmaker and son of the photographer Herman Matter and painter Mercedes Matter, announced the discovery of 32 early Jackson Pollock drip paintings. The full story sounded a lot like a New York version of Michael Frayn's novel *Headlong*. Matter's parents were friends of Pollock, close enough to the artist that photographs exist of them lounging about with Pollock. At least one biography of Pollock mentions the purchase of some small Pollock works by the Matters, and it is reported that Herman Matter's signature is on the back of these works, labeling them as Pollock originals. It seems that the paintings had been languishing in a metal storage bin out on Long Island until they were discovered by Mr. Matter, not far from Pollock's former home in the Hamptons and the scene of his fatal car accident. The press releases were accompanied by pictures of a grinning Alex Matter holding one of these paintings, which, at a distance and at the resolution of a newspaper photograph, looked suitably spattered. It's a good story and certainly a plausible one.

But, in fact, hardly a year goes by in which someone doesn't claim to have found a long-lost Pollock—that's what Richard Taylor, a consultant for the Pollock-Krasner Foundation tells me. Most of these claims are disposed of easily, their faults ranging from the obvious (misspelled signatures—"Jackson Pollack") to the subtle (materials that were only available after Pollock's death). Some fakes are veterans of the auction circuit and reappear every few years with a new story ("Some hitchhiker gave this to a truck driver, who brought it into my gallery…"), while other schemes are almost as ingenious as the artist himself, even going to the extremes of inventing a provenance by

inserting appropriately faded pictures of the "discovered" work into obscure hard-to-find gallery catalogs.

So, how to tell real from fake? In spite of the fact that there is only one right answer, generally, evaluating claims of authenticity in the world of modern art is, well, more art than science. Slam-dunks are the situations in which the work is accompanied by an airtight provenance, consisting of documentation of every single person or institution that has ever owned the work (the back of a painting can look like a well-traveled steamer trunk, covered with labels marking its appearance in exhibitions). Such works are also accompanied by records of any conservation incurred and a list of events detailing the touch-up here or the cleaning there. Failing this sort of complete biography, the truth begins to get a little murkier. Sometimes science can help. X-rays can reveal underpaintings or preliminary sketches. Analysis of the materials provides clues, including information as to the age of the work. But even this hard evidence can go only so far. Several years ago the photography market was flooded with "newly discovered" Man Ray prints, their authenticity supported by the fact that the photographic paper was exactly of the kind that Man Ray was known to have used 60 years ago. It was only later discovered that the forger had managed to find an old cache of this paper, obtain access to Man Ray's negatives, and use this to make a new set of prints.

Watery Paths (1947) by Jackson Pollock. From the Galleria Nazionale d'Arte Moderna, Rome, Italy.

So, enter the human actor. Connoisseurs steeped in the work of the artist in question apply the "sniff test" and either the light bulb goes on or it doesn't. And the fact is that different connoisseurs of the same artist can have different senses of smell. Ultimately, in the case of Alex Matter's bundle of proposed Pollocks, the experts are taking a close look at each painting and answering the question, "Is this work characteristic of the style of Jackson Pollock at this stage in his career?" This is the question that Richard Taylor asked when the Pollock-Krasner Foundation called him in to evaluate the Matter cache—but the way in which he

Peasant Dance by Pieter Bruegel the Elder. Oil on oakwood (1568).

ABOVE: ERICH LESSING/ART RESOURCE, NY

differs from his connoisseur colleagues is that he answered the question with mathematics. Taylor seems to have found a way to extract a numerical signature that identifies a work as an original Pollock. More precisely, Taylor has determined that a statistical analysis of the numbers that comprise the digital representation of the work can reliably separate authentic Pollocks from fakes. In so doing, Taylor's work is paving the way for a new world of visual "stylometry," that is, a growing discipline that is making math out of that seemingly impossible to define thing that is style. It is a subject that actually goes back to the mid-1800s, but has recently exploded in our world in which all actions and creations are fodder for the computer and for those with broad vision. The tools of trade find their way into all aspects of our life, helping to distill not only styles of artistic creation, but styles of all sorts of actions.

MEASURING STYLE

Mathematical analysis or statistical analysis requires that something be measured, so it's not surprising that as applied to the arts it is able to first find a foothold in literature. Books, essays, any sort of written output,

present a numerical collection, containing all kinds of primary data. Letters, words, sentences and so on, can be counted and basic statistics can be gathered: How often is a given word used? What is the distribution of word length? What is the distribution of sentence length? etc. Do this and you are guaranteed to get numbers. What is not guaranteed, and hence is surprising, is that out of these numbers, patterns emerge, both on the scale of society, as well as the individual. In the former we see the patterns that seem to be intrinsic to any form of communication, and in the latter, we seem to be able to distill aspects of the idiosyncratic patterns of usage that form the basis of a person's writing style.

The idea that mathematics might be useful for determining authorship is usually attributed to the 19th-century British mathematician Augustus de Morgan. De Morgan was in many respects ahead of his time, especially as regards to what appears to have been a broad-based investigation of the power and possibility of the formalization of thought. An interest in the formal or quantifiable aspects of creative work easily fits into this program.

As recorded in his wife's memoirs, de Morgan wrote a letter to a friend in 1854 that states, "It has always run in my head that a little expenditure of money would settle questions of authorship" by determining if the writings of "the latter do not deal in longer words than the writings of the former." It must have been the case that de Morgan never had the extra pocket change to pursue this thought, as it was several years later in 1886 that the first published account of a mathematical analysis of literature appears, written by American physicist Thomas Mendenhall. The subject is Shakespeare, a favored focus of the question of attribution. Mendenhall tries to distinguish between Francis Bacon and William Shakespeare according to their relative use of four-letter words. It's not a success, but nevertheless, a discipline is born. It finds a name when 10 years later, Wincenty Lutoslawski looks at 500 numerical attributes in each of Plato's dialogues in order to reconstruct the order in which they were produced, working with the basic philosophy that works that are close mathematically should be close temporally. He called his methodology "stylometry."

> Taylor's work is paving the way for a new world of visual "stylometry," that is, a growing discipline that is making math out of that seemingly impossible to define thing that is style.

Modern trends in stylometry pull from the full bag of tricks of artificial intelligence and advanced statistical analysis. Some approaches focus on aspects of predictability—using the empirical likelihood that one word is followed by another. These ideas were first proposed in the early 1900s by the Russian mathematician A. A. Markov who used them to construct a very simple model for the cadences found in Pushkin's poem "Eugene Onegin." Today, so-called "Markov chains" are among the most commonly used tools in the mathematical modeler's workbelt and can be found identifying patterns in all sorts of places, ranging from genetic to financial data.

Word frequencies used in one way or another remain the heart and soul of literary stylometry. Some of the most successful techniques focus on usage statistics of "function words." As opposed to "content words," these include pronouns, conjunctions, and prepositions, which generally carry very little non-contextual meaning and serve instead as grammatical connective glue. Literary style, it appears, resides in the degree to which we choose "that" rather than "which" or "however" as opposed to "nevertheless." The starting material for many a stylometric analysis begins by first isolating and recording the frequencies of the favorite function words among works of known authorship and then, in one way or another, considering the degree to which the frequency pattern in a contested

work is statistically similar to the patterns in the secure works. Among other examples, function word usage has been used to distinguish between the writings of Alexander Hamilton and James Madison as well as to pinpoint authorship in the *Wizard of Oz* series.

One of the most striking results in the field is the discovery that there are certain patterns of usage that seem to be simply intrinsic to the act of communication. In 1949, using a corpus of a range of works by a number of different authors, Harvard linguist George Zipf discovered a remarkable empirical fact, known today as "Zipf's law." It states that the result of multiplying a given word frequency by its "rank" (the most frequent word has rank one, the second most frequent is rank two, and so on) was approximately the same over all the words in the corpus. Zipf published his analysis in an amazing book titled *Human Behavior and the Principle of Least Effort,* which derives its name from the basic argument that predicts the Zipfian discovery: Imagine any author as a chance-driven machine in which at each step a coin is tossed—if it lands heads up, then a previously used word is chosen at random, while if it lands tails up, then a new word is written down. This is a "rich get richer" sort of model in which the more a word is used in the past, the more likely it is that it will attract more use in the future. In fact, Zipf finds similar relations (called "power laws"—a familiar distribution in the complex systems world) among all sorts of ranked lists, ranging from sizes of towns (here you would consider the product of the population of the town with its rank) to income distributions (where it also goes by the name of "Pareto's law").

Words can be counted—that's the main reason that stylometry came first to literature. But, other art forms have natural numbers too. Given the success of literary stylometry and the empirical ubiquity of Zipfian behavior, it's something of a surprise that it was not until just a few years ago that a broad analysis of that other great symbolic language that is musical composition was undertaken. Charleston College computer scientist Bill Manaris led a small group of researchers that counted note usage over a range of composers and works, and discovered a basic power law structure. Using that as the foundation of his analysis, he and his colleagues were able to derive a collection of statistical features from musical scores that successfully allow an automatic classification of musical works from jazz, classical, and rock n' roll.

ACTIONS SPEAK AS LOUDLY AS WORDS

Even with movement, it is possible to distill style from the numbers. In its measurement of angular displacements of joints and relative displacements of limbs, kinematic analysis reduces human movement to streams of numbers. "Plug n' play" animation software is evidence of the fact that there is a basic mathematical formulation to the way in which we move—there is an average walk, run, jump, and so on. Stylometry comes into the picture as a means to give these virtual folks the style of movement of particular real people, accomplished by motion-capture systems ("mocap") that can track and record the movements of a collection of sensors or transmitters worn about the body. The acquired data can provide a numerical record of the particular way that a person walks or sips tea—and then the animator can take that away, using the manner in which that person differs from the average to give a personal touch to an animated avatar. Motion capture is a modern updating of the way in which the early Disney animators worked—often tracing over stills gleaned from footage of the movements of professional dancers and clowns in order to acquire a feel for a basic movement style. Mocap is now a standard Hollywood technique, responsible for personified performances such as Tom Hanks's animated turn as the conductor in *Polar Express*. This Frankensteinian "retargeting" of the motions of the living to the lifeless has even been attempted from cartoon to cartoon. Through an adaptation of the mathematics of mocap, Fred Flintstone can take dance lessons from Mickey Mouse.

The success of motion capture for animation is evidence that actions do indeed speak at least as loudly as words. This is an idea that is driving a new generation of marketing and

advertising. Your shopping has a style that is encoded in the trace of clicks and eyeball dwells that you leave on the web, your check-out list at the grocery store, and your monthly credit card bill. The electronic hectorings "If you liked this, then you'll like that!" are mathematical statements based on a geometry of sales-space—an item that you buy is encoded as a list of numerical attributes, and like the two-number lists that make up the x, y coordinates studied in a high school geometry class, these lists of item coordinates give geometric meaning to your buying habits. If marketers find you shopping in one region of their abstract product space, they've a pretty good idea that you may like some items nearby that their specially designed math marketing goggles allow them to see. This approach can even be applied to shopping for a mate—that's at least what some of the online dating services rely on—your personal style has a shape, in the most mathematical of terms.

PAINTING BY NUMBERS

That the visual arts have been among the last to embrace any form of stylometry is perhaps more a matter of tradition than anything else. As a discipline, connoisseurship has scientific origins, usually attributed to Giovanni Morelli, a 19th-century Swiss–Italian government official with a deep appreciation for the arts. It was in a large part shaped by an early education focused on the sciences, and, in particular, his experiences

The crags of a mountain range are replicated in the nooks and crannies of the stones that comprise them, or the eddies of a turbulent river flow are themselves composed of eddies within eddies, within eddies.

accompanying the renowned and pioneering paleontologist and naturalist Louis Agassiz on his glacier expeditions in Switzerland.

Morelli brought the skills of an expert naturalist to the problem of looking at, comparing, and, finally, classifying works of art. In his major lifework, *Italian Painters: Critical Studies of their Works*, Morelli foreshadows the still unborn science of literary styometry as he writes, "As most men, both speakers and writers, make use of habitual modes of expression, favorite words and sayings, which they often employ involuntarily and sometimes even most appropriately, so almost every painter has his own peculiarities, which escape him without being aware of it." According to Morelli, these "peculiarities" would find expression in the quiet corners of a work of art. Thus the "Morellian method" relies on the comparison of seemingly minor details in paintings: folds in drapery, a fingernail, or an earlobe. In Morelli's view, it was only in details such as these, which he called the "*grundformen*" (fundamental forms) of the artist, that the forces of tradition or schooling would be diminished enough so that the artist's true nature could shine through. In spirit, it is a visual form of Zipf's principle of least effort, a view of the artistic output as shaped by a battle between the expectations of the receiver and the predilections of the sender.

Morelli was only able to apply his ideas with the tools he had at hand, mainly his eyes. Lacking digital scanners and image processing software, he built an internal database of *grundformen* that, to some degree, would permit him to distinguish between details such as hands painted by Boticelli and Bellini. But with today's technology, we can begin to make mathematics out of Morelli. Through the use of "wavelet analysis," a mathematical technique originally developed in the 1980s as a means of determining within sonar data the fingerprint of oil deposits beneath the ocean floor, my Dartmouth colleagues and I have made headway. We have found that a statistical summary of the density of simple linear elements in some of the drawings of the great Flemish

artist Pieter Bruegel the Elder, as extracted from high resolution digital scans of the originals, can provide a numerical signature that seems to act as a classifier for Bruegel's work. Even more Morellian than that, we have applied the same technique to the comparison of details within *Madonna and Child,* a huge altarpiece attributed to the great Renaissance master Perugino, in the hopes of determining the number of artists who contributed to the work. Whether or not the exploration for oil in the oceans has resulted in a definitive tool for the exploration of oils on the wall is still up for debate.

Wavelets are an example of a multiscale analysis—that is, they proceed by analyzing a work at successive levels of detail, like examining a work through varying the magnification on a microscope. In a wavelet analysis, what becomes important is the difference between what is measured at two scales, in a sense, representing the original image as some baseline structure to which is added a successive layering of detail.

AUTHENTICATION THROUGH FRACTALS

Multiscale analysis is also at the heart of Richard Taylor's approach to Pollock. Taylor's investigations have not been driven by questions of provenance (although due to the frequent requests for authentication he has set-up a non-profit company to manage this work as well as to protect himself in the case of lawsuits). He is mainly interested in the work, as both a scientist and an artist. He had taken a year's leave at one point in his career to devote himself to his own painting, but after a year decided that he was better off not quitting his day job. He brings a physicist's eye to the arts, and in the case of Pollock, it has been a perfect storm of art and science that has enabled Taylor to find his own research in the works of this abstract expressionist master—the "chaos" for which one critic famously denounced Pollock's work in the 1950s is something that Taylor saw quite literally as the mathematics of fractal geometry.

When Pollock so famously said, "I am nature," or that "My concern is with the rhythm of nature…the way the ocean moves," he was possibly closer to the truth than anyone gave him credit for and probably closer than he knew himself.

The word "fractal" was coined in 1967 by the IBM mathematician Benoit Mandelbrot to encompass the geometric character of natural objects. The perfect lines, planes, and spheres of Euclidean geometry are Platonic abstractions, good for a first approximation to things like coastlines, landscapes, and clouds, but clearly fall short at describing the variation of the natural world. Mandelbrot noticed that the character of such natural phenomena was a similarity in scale—that at each increase in magnification, the structures of nature, complicated though they are ("fractal" is derived from the same root as fragment and fracture), repeat themselves, maybe not precisely, but to a degree that can be quantified. The crags of a mountain range are replicated in the nooks and crannies of the stones that comprise them, or the eddies of a turbulent river flow are themselves composed of eddies within eddies, within eddies. This is a piece of the connection between chaos and fractal geometry—the chaos that we now know colloquially in the metaphor of the hurricane in Texas generated by the flapping wings of the butterfly over China is a phenomenon that when put into mathematical pictures (not unlike those that can be seen on The Weather Channel) give images that exhibit this sort of self-similarity.

A famous example of a natural fractal is the irregular outline of the coast of England, which is, to a degree, replicated in any stretch of shore beneath the cliffs of Dover. The latter example, due to the British mathematician and polymath, Lewis Fry Richardson, was Mandelbrot's inspiration for the quantification of this irregularity in terms of its "fractal dimension," a

number that effectively measures the complexity of a shape in terms of the degree to which it fills space at a given scale. The crinklier a line is, the more space it occupies in a box that surrounds it. Now, imagine a shape where as you crank up the magnification, that sort of "misbehavior" is replicated: You've got yourself a true fractal.

A perfectly straight line has fractal dimension equal to one, while a square region has fractal dimension equal to two. Nature is generally somewhere in between: the coastline of England, the waves within waves of a stormy sea, the branches within branches of a fern leaf, our own circulatory or pulmonary system, or as it turns out, the skeins of paint in a Pollock drip painting. That is Taylor's discovery. When Pollock so famously said, "I am nature," or that "My concern is with the rhythm of nature…the way the ocean moves," he was possibly closer to the truth than anyone gave him credit for and probably closer than he knew himself.

Taylor has examined many of Pollock's works and found a remarkable degree of regularity in the fractal dimension that can be computed by examining different color layers in the paintings. Firstly, what is remarkable is that Pollock could regularly achieve fractal structure. Taylor's personal attempts at such a result were only successful when he came up with the idea of hanging a bed sheet from a tree and allowing the measurable fractal nature of the wind to be realized in dripped paint blown onto the sheet. Even more, there appears to be a fractal dimension to Pollock's work that is characteristic of a given period, so that Pollock did, over periods of time, reliably reproduce in his work a small range of fractal dimensions. In fact, Taylor claims even more, that in his examinations he finds evidence for two distinct fractal dimensions as might be predicted by a documented two-step working style in which Pollock would lay down a broad underlayer to which he would later add detail.

When presented with a would-be Pollock, Taylor performs the digital analysis and checks to see if the numbers jibe with those that have been computed for Pollock's known work of a given period. Taylor's analysis of the Matter collection suggested that the drip paintings were forgeries. However, in some related work, John Elton and Yang Wang of Georgia Tech; Jim Coddington, Chief Conservator at New York's Museum of Modern Art; and I have determined that a generalization of fractal dimension, called *multifractal* analysis, may provide a more textured signature for the work.

It's significant that Taylor found a digital signature for Pollock. But what might be even more significant is that the art world paid attention to it, for this shows the art/science boundaries are continuing to become fuzzier and fuzzier. Presumably, this is just the beginning, although there will surely be artists whose work defeats a statistical approach. In the spring of 2007, five teams of researchers will converge on the Van Gogh Museum in Amsterdam to present the results of a year-long study aimed at uncovering a digital signature for Vincent van Gogh. And other methods of this type of work are emerging. An interesting and very general approach to finding a style in any digital media is work of the Dutch information theorist Paul Vitanyi, whose analysis focuses on the information content (in a statistical sense) of the work. His media-free approach is one that allows any collection of numbers to be compared to any other, making possible the idea of comparing works of art to works of literature.

Stylometry opens us up to a world in which we are defined by our digital trail—the words we write, the websites we visit, the pictures we store, summarized in a statistical fingerprint. We are our actions. How very existential. ◀

Dan Rockmore is professor of mathematics and computer science at Dartmouth College and a member of the SFI External Faculty.