Bi-weekly Random Bits from the Internet

2015-02-14

(I WATCHED ONE OF LOUIS' SHOW WHEN HE WAS STILL MARRIED AND THAT IS A TERRIBLE THING TO SAY)

The Pursuit of Beauty

Alec Wilkinson, New Yorker Feb 02, 2015 Issue

I don't see what difference it can make now to reveal that I passed high-school math only because I cheated. I could add and subtract and multiply and divide, but I entered the wilderness when words became equations and x's and y's. On test days, I sat next to Bob Isner or Bruce Gelfand or Ted Chapman or Donny Chamberlain—smart boys whose handwriting I could read—and divided my attention between his

desk and the teacher's eyes. Having skipped me, the talent for math concentrated extravagantly in one of my nieces, Amie Wilkinson, a professor at the University of Chicago. From Amie I first heard about Yitang Zhang, a solitary, part-time calculus teacher at the University of New Hampshire who received several prizes, including a MacArthur award in September, for solving a problem that had been open for more than a hundred and fifty years.

The problem that Zhang chose, in 2010, is from number theory, a branch of pure mathematics. Pure mathematics, as opposed to applied mathematics, is done with no practical purposes in mind. It is as close to art and philosophy as it is to engineering. "My result is useless for industry," Zhang said. The British mathematician G. H. Hardy wrote in 1940 that mathematics is, of "all the arts and sciences, the most austere and the most remote." Bertrand Russell called it a refuge from "the dreary exile of the actual world." Hardy believed emphatically in the precise aesthetics of math. A mathematical proof, such as Zhang produced, "should resemble a simple and clear-cut constellation," he wrote, "not a scattered cluster in the Milky Way." Edward Frenkel, a math professor at the University of California, Berkeley, says Zhang's proof has "a renaissance beauty," meaning that though it is deeply complex, its outlines are easily apprehended. The pursuit of beauty in pure mathematics is a tenet. Last year, neuroscientists in Great Britain discovered that the same part of the brain that is activated by art and music was activated in the brains of mathematicians when they looked at math they regarded as beautiful.

Zhang's problem is often called "bound gaps." It concerns prime numbers—those which can be divided cleanly only by one and by themselves: two, three, five, seven, and so on—and the question of whether there is a boundary within which, on an infinite number of occasions, two consecutive prime numbers can be found, especially out in the region where the numbers are so large that it would take a book to print a single one of them. Daniel Goldston, a professor at San Jose State University; János Pintz, a fellow at the Alfréd Rényi Institute of Mathematics, in Budapest; and Cem Yıldırım, of Boğaziçi University, in Istanbul, working together in 2005, had come closer than anyone else to establishing whether there might be a boundary, and what it might be. Goldston didn't think he'd see the answer in his lifetime. "I thought it was impossible," he told me.

Zhang, who also calls himself Tom, had published only one paper, to quiet acclaim, in 2001. In 2010, he was fifty-five. "No mathematician should ever allow himself to forget that mathematics, more than any other art or science, is a young man's game," Hardy wrote. He also wrote, "I do not know of an instance of a major mathematical advance initiated by a man past fifty." Zhang had received a Ph.D. in algebraic geometry from Purdue in 1991. His adviser, T. T. Moh, with whom he parted unhappily, recently wrote a description on his Web site of Zhang as a graduate student: "When I looked into his eyes, I found a disturbing soul, a burning bush, an explorer who wanted to reach the North Pole." Zhang left Purdue without Moh's support, and, having published no papers, was unable to find an academic job. He lived, sometimes with friends, in Lexington, Kentucky, where he had occasional work, and in New York City, where he also had friends and occasional work. In Kentucky, he became involved with a group interested in Chinese democracy. Its slogan was "Freedom, Democracy, Rule of Law, and Pluralism." A member of the group, a chemist in a lab, opened a Subway franchise as a means of raising money. "Since Tom was a genius at numbers," another member of the group told me, "he was invited to help him." Zhang kept the books. "Sometimes, if it was busy at the store, I helped with the cash register," Zhang told me recently. "Even I knew how to make the sandwiches, but I didn't do it so much." When Zhang wasn't working, he would go to the library at the University of Kentucky and read journals in algebraic geometry and number theory. "For years, I didn't really keep up my dream in mathematics," he said.

"You must have been unhappy."

He shrugged. "My life is not always easy," he said.

With a friend's help, Zhang eventually got his position in New Hampshire, in 1999. Having chosen bound gaps in 2010, he was uncertain of how to find a way into the problem. "I am thinking, Where is the door?" Zhang said. "In the history of this problem, many mathematicians believed that there should be a door, but they couldn't find it. I tried several doors. Then I start to worry a little that there is no door."

"Were you ever frustrated?"

"I was tired," he said. "But many times I just feel peaceful. I like to walk and think. This is my way. My wife would see me and say, 'What are you doing?' I said, 'I'm working, I'm thinking.' She didn't understand. She said, 'What do you mean?' " The problem was so complicated, he said, that "I had no way to tell her."

According to Deane Yang, a professor of mathematics at the New York University Polytechnic School of Engineering, a mathematician at the beginning of a difficult problem is "trying to maneuver his way into a maze. When you try to prove a theorem, you can almost be totally lost to knowing exactly where you want to go. Often, when you find your way, it happens in a moment, then you live to do it again."

Zhang is deeply reticent, and his manner is formal and elaborately polite. Recently, when we were walking, he said, "May I use these?" He meant a pair of clip-on shades, which he held toward me as if I might want to examine them first. His enthusiasm for answering questions about himself and his work is slight. About half an hour after I had met him for the first time, he said, "I have a question." We had been talking about his childhood. He said, "How many more questions you going to have?" He depends heavily on three responses: "Maybe," "Not so much," and "Maybe not so much." From diffidence, he often says "we" instead of "I," as in, "We may not think this approach is so important." Occasionally, preparing to speak, he hums. After he published his result, he was invited to spend six months at the Institute for Advanced Study, in Princeton. The filmmaker George Csicsery has made a documentary about Zhang, called "Counting from Infinity," for the Mathematical Sciences Research Institute, in Berkeley, California. In it, Peter Sarnak, a member of the Institute for Advanced Study, says that one day he ran into Zhang and said hello, and Zhang said hello, then Zhang said that it was the first word he'd spoken to anyone in ten days. Sarnak thought that was excessive, even for a mathematician, and he invited Zhang to have lunch once a week.

Matthew Emerton, a professor of math at the University of Chicago, also met Zhang at Princeton. "I wouldn't say he was a standard person," Emerton told me. "He wasn't gregarious. I got the impression of him being reasonably internal. He had received another prize, so the people around him were talking about that. Probably most mathematicians are very low-key about getting a prize, because you're not in it for the prize, but he seemed particularly low-key. It didn't seem to affect him at all." Deane Yang attended three lectures that Zhang gave at Columbia in 2013. "You expect a guy like that to want to show off or explain how smart he is," Yang said. "He gave beautiful lectures, where he wasn't trying to show off at all." The first talk that Zhang gave on his result was at Harvard, before the result was published. A professor there, Shing-Tung Yau, heard about Zhang's paper, and invited him. About fifty people showed up. One of them, a Harvard math professor, thought Zhang's talk was "pretty incomprehensible." He added, "The problem is that this stuff is hard to talk about, because everything hinges on some delicate technical understandings." Another Harvard professor, Barry Mazur, told me that he was "moved by his intensity and how brave and independent he seemed to be."

In New Hampshire, Zhang works in an office on the third floor of the math and computer-science building. His office has a desk, a computer, two chairs, a whiteboard, and some bookshelves. Through a window he looks into the branches of an oak tree. The books on his shelves have titles such as "An Introduction to Hilbert Space" and "Elliptic Curves, Modular Forms, and Fermat's Last Theorem." There are also books on modern history and on Napoleon, who fascinates him, and copies of Shakespeare, which he reads in Chinese, because it's easier than Elizabethan English.

Eric Grinberg, the chairman of the math department at the University of Massachusetts Boston, was a colleague of Zhang's in New Hampshire from 2003 to 2010. "Tom was very modest, very unassuming, never asked for anything," Grinberg told me. "We knew he was working on something important. He uses paper and a pencil, but the only copy was on his computer, and about once a month I would go in and ask, 'Do you mind if I make a backup?' Of course, it's all in his head anyway. He's above average in that."

Zhang's memory is abnormally retentive. A friend of his named Jacob Chi said, "I take him to a party sometimes. He doesn't talk, he's absorbing everybody. I say, 'There's a human decency; you must talk to people, please.' He says, 'I enjoy your conversation.' Six months later, he can say who sat where and who started a conversation, and he can repeat what they said."

"I may think socializing is a way to waste time," Zhang says. "Also, maybe I'm a little shy."

A few years ago, Zhang sold his car, because he didn't really use it. He rents an apartment about four miles from campus and rides to and from his office with students on a school shuttle. He says that he sits on the bus and thinks. Seven days a week, he arrives at his office around eight or nine and stays until six or seven. The longest he has taken off from thinking is two weeks. Sometimes he wakes in the morning thinking of a math problem he had been considering when he fell asleep. Outside his office is a long corridor that he likes to walk up and down. Otherwise, he walks outside.

Zhang met his wife, to whom he has been married for twelve years, at a Chinese restaurant on Long Island, where she was a waitress. Her name is Yaling, but she calls herself Helen. A friend who knew them both took Zhang to the restaurant and pointed her out. "He asked, 'What do you think of this girl?' " Zhang said. Meanwhile, she was considering him. To court her, Zhang went to New York every weekend for several months. The following summer, she came to New Hampshire. She didn't like the winters, though, and moved to California, where she works at a beauty salon. She and Zhang have a house in San Jose, and he spends school vacations there.

Until Zhang was promoted to professor, last year, as a consequence of his proof, his appointment had been tenuous. "I was chair of the math department, and I had to go to him from time to time and remind him this was not a permanent position," Eric Grinberg said. "We were grateful to him, but it's not guaranteed. He always said that he very much appreciated the time he had spent in New Hampshire."

Zhang devoted himself to bound gaps for a couple of years without finding a door. "We couldn't see any hope," he said. Then, on July 3, 2012, in the middle of the afternoon, "within five or ten minutes, the way is open."

Zhang was in Pueblo, Colorado, visiting his friend Jacob Chi, who is a music professor at Colorado State University-Pueblo. A few months earlier, Chi had reminded Zhang that he had promised one day to teach his son, Julius, calculus, and since Julius was about to be a senior in high school Chi had called and asked, "Do you keep your promise?" Zhang spent a month at the Chis'. Each morning, he and Julius worked for about an hour. "He didn't have a set curriculum," Julius told me. "It all just flowed from his memory. He mentioned once that he didn't have any numbers in his phone book. He memorized them all."

Zhang had planned a break from work in Colorado, and hadn't brought any notes with him. On July 3rd, he was walking around the Chis' back yard. "We live in the mountains, and the deer come out, and he was smoking a cigarette and watching for the deer," Chi said. "No deer came," Zhang said. "Just walking and thinking, this is my way." For about half an hour, he walked around at a loss.

In "The Psychology of Invention in the Mathematical Field," published in 1945, Jacques Hadamard quotes a mathematician who says, "It often seems to me, especially when I am alone, that I find myself in another world. Ideas of numbers seem to live. Suddenly, questions of any kind rise before my eyes with their answers." In the back yard, Zhang had a similar experience. "I see numbers, equations, and something even—it's hard to say what it is," Zhang said. "Something very special. Maybe numbers, maybe equations—a mystery, maybe a vision. I knew that, even though there were many details to fill in, we should have a proof. Then I went back to the house."

Zhang didn't say anything to Chi about his breakthrough. That evening, Chi was conducting a dress rehearsal for a Fourth of July concert in Pueblo, and Zhang went with him. "After the concert, he couldn't stop humming 'The Stars and Stripes Forever,' " Chi said. "All he would say was 'What a great song.' "

I asked Zhang, "Are you very smart?" and he said, "Maybe, a little." He was born in Shanghai in 1955. His mother was a secretary in a government office, and his father was a college professor whose field was electrical engineering. As a small boy, he began "trying to know everything in mathematics," he said. "I became very thirsty for math." His parents moved to Beijing for work, and Zhang remained in Shanghai with his grandmother. The revolution had closed the schools. He spent most of his time reading math books that he ordered from a bookstore for less than a dollar. He was fond of a series whose title he translates as "A Hundred Thousand Questions Why." There were volumes for physics, chemistry, biology, and math. When he didn't understand something, he said, "I tried to solve the problem myself, because no one could help me."

Zhang moved to Beijing when he was thirteen, and when he was fifteen he was sent with his mother to the countryside, to a farm, where they grew vegetables. His father was sent to a farm in another part of the country. If Zhang was seen reading books on the farm, he was told to stop. "People did not think that math was important to the class struggle," he said. After a few years, he returned to Beijing, where he got a job in a factory making locks. He began studying to take the entrance exam for Peking University, China's most respected school: "I spent several months to learn all the high-school physics and chemistry, and several to learn history. It was a little hurried." He was admitted when he was twenty-three. "The first year, we studied calculus and linear algebra—it was very exciting," Zhang said. "In the last year, I selected number theory as my specialty." Zhang's professor insisted, though, that he change his major to algebraic geometry, his own field. "I studied it, but I didn't really like it," Zhang said. "That time in China, still the idea was like this: the individual has to follow the interest of the whole group, the country. He thought algebraic geometry was more important than number theory. He forced me. He was the university president, so he had the authority."

During the summer of 1984, T. T. Moh visited Peking University from Purdue and invited Zhang and several other students, recommended to him by Chinese professors, to do graduate work in his department. One of Moh's specialties is the Jacobian conjecture, and Zhang was eager to work on it. The Jacobian conjecture, a problem in algebraic geometry that was introduced in 1939 and is still unsolved, stipulates certain simple conditions that, if satisfied, enable someone to solve a series of complicated equations. It is acknowledged as being beyond the capacities of a graduate student and approachable by only the most accomplished algebraic geometrists. A mathematician described it to me as a "disaster problem," for the trouble it has caused. For his thesis, Zhang submitted a weak form of the conjecture, meaning that he attempted to prove something implied by the conjecture, rather than to prove the conjecture itself.

After Zhang received his doctorate, he told Moh that he was returning to number theory. "I was not the happiest," Moh wrote me. "However, I was for the student's right to change fields, so I kept my smile and said bye to him. For the past 22 years, I knew nothing about him."

After graduating, most of the Chinese students went into either computer science or finance. One of them, Perry Tang, who had known Zhang in China, took a job at Intel. In 1999, he called Zhang. "I thought it was unfair for him not to have a professional job," Tang said. He and Zhang had a classmate at Peking University who had become a professor of math at the University of New Hampshire, and when the friend said that he was looking for someone to teach calculus Tang recommended

Zhang. “He decided to try him at a temporary position,” Tang said.

Zhang finished “Bounded Gaps Between Primes” in late 2012; then he spent a few months methodically checking each step, which he said was “very boring.” On April 17, 2013, without telling anyone, he sent the paper to Annals of Mathematics, widely regarded as the profession’s most prestigious journal. In the Annals archives are unpublished papers claiming to have solved practically every math problem that anyone has ever thought of, and others that don’t really exist. Some are from people who “know a lot of math, then they go insane,” a mathematician told me. Such people often claim that everyone else who has attacked the problem is wrong. Or they announce that they have solved several problems at once, or “they say they have solved a famous problem along with some unified-field theory in physics,” the mathematician said. Journals such as Annals are always skeptical of work from someone they have never heard of.

In 2013, Annals received nine hundred and fifteen papers and accepted thirty-seven. The wait between acceptance and publication is typically around a year. When a paper arrives, “it is read quickly, for worthiness,” Nicholas Katz, the Princeton professor who is the journal’s editor, told me, and then there is a deep reading that can take months. “The paper I can’t evaluate off the top of my head, my role is to know whom to ask,” Katz said. “In this case, the person wrote back pretty quickly to say, ‘If this is correct, it’s really fantastic. But you should be careful. This guy posted a paper once, and it was wrong. He never published it, but he didn’t take it down, either.’ ” The reader meant a paper that Zhang posted on the Web site arxiv.org, where mathematicians often post results before submitting them to a journal, in order to have them seen quickly. Zhang posted a paper in 2007 that fell short of a proof. It concerned another famous problem, the Landau-Siegel zeros conjecture, and he left it up because he hopes to correct it.

Katz sent “Bounded Gaps Between Primes” to a pair of readers, who are called referees. One of them was Henryk Iwaniec, a professor at Rutgers, whose work was among that which Zhang had drawn on. “I glanced for a few minutes,” Iwaniec told me. “My first impression was: So many claims have become wrong. And I thought, I have other work to do. Maybe I’ll postpone it. Remember that he was an unknown guy. Then I got a phone call from a friend, and it happened he was also reading the paper. We were going to be together for a week at the Institute for Advanced Study, and the intention was to do other work, but we were interrupted with this paper to read.”

Iwaniec and his friend, John Friedlander, a professor at the University of Toronto, read with increasing attention. “In these cases, you don’t read A to Z,” Iwaniec said.

“You look first at where is the idea. There had been nothing written on the subject since 2005. The problem was too difficult to solve. As we read more and more, the chance that the work was correct was becoming really great. Maybe two days later, we started looking for completeness, for connections. A few days passed, we’re checking line by line. The job is no longer to say the work is fine. We are looking to see if the paper is truly correct.”

After a few weeks, Iwaniec and Friedlander wrote to Katz, “We have completed our study of the paper ‘Bounded Gaps Between Primes’ by Yitang Zhang.” They went on, “The main results are of the first rank. The author has succeeded to prove a landmark theorem in the distribution of prime numbers.” And, “Although we studied the arguments very thoroughly, we found it very difficult to spot even the smallest slip. . . . We are very happy to strongly recommend acceptance of the paper for publication in the Annals.”

Once Zhang heard from Annals, he called his wife in San Jose. “I say, ‘Pay attention to the media and newspapers,’ ” he said. “ ‘You may see my name,’ and she said, ‘Are you drunk?’ ”

No formula predicts the occurrence of primes—they behave as if they appear randomly. Euclid proved, in 300 B.C., that there is an infinite number of primes. If you imagine a line of all the numbers there are, with ordinary numbers in green and prime numbers in red, there are many red numbers at the beginning of the line: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, and 47 are the primes below fifty. There are twenty-five primes between one and a hundred; 168 between one and a thousand; and 78,498 between one and a million. As the primes get larger, they grow scarcer and the distances between them, the gaps, grow wider.

Prime numbers have so many novel qualities, and are so enigmatic, that mathematicians have grown fetishistic about them. Twin primes are two apart. Cousin primes are four apart, sexy primes are six apart, and neighbor primes are adjacent at some greater remove. From “Prime Curios!,” by Chris Caldwell and G. L. Honaker, Jr., I know that an absolute prime is prime regardless of how its digits are arranged: 199; 919; 991. A beastly prime has 666 in the center. The number 700666007 is a beastly palindromic prime, since it reads the same forward and backward. A circular prime is prime through all its cycles or formulations: 1193, 1931, 9311, 3119. There are Cuban primes, Cullen primes, and curved-digit primes, which have only curved numerals—0, 6, 8, and 9. A prime from which you can remove numbers and still have a prime is a deletable prime, such as 1987. An emirp is prime even when you reverse it: 389, 983. Gigantic primes have more than ten thousand digits, and holey primes have only digits with holes (0, 4, 6, 8, and 9). There are Mersenne primes; minimal primes; naughty primes, which are made mostly from zeros (naughts); ordinary primes; Pierpont primes; plateau primes, which have the same interior numbers and smaller numbers on the ends, such as 1777771; snowball primes, which are prime even if you haven't finished writing all the digits, like 73939133; Titanic primes; Wagstaff primes; Wall-Sun-Sun primes; Wolstenholme primes; Woodall primes; and Yarborough primes, which have neither a 0 nor a 1.

"Bounded Gaps Between Primes" is a back-door attack on the twin-prime conjecture, which was proposed in the nineteenth century and says that, no matter how far you travel on the number line, even as the gap widens between primes you will always encounter a pair of primes that are separated by two. The twin-prime conjecture is still unsolved. Euclid's proof established that there will always be primes, but it says nothing about how far apart any two might be. Zhang established that there is a distance within which, on an infinite number of occasions, there will always be two primes.

"You have to imagine this coming from nothing," Eric Grinberg said. "We simply didn't know. It is like thinking that the universe is infinite, unbounded, and finding it has an end somewhere." Picture it as a ruler that might be applied to the line of green and red numbers. Zhang chose a ruler of a length of seventy million, because a number that large made it easier to prove his conjecture. (If he had been able to prove the twin-prime conjecture, the number for the ruler would have been two.) This ruler can be moved along the line of numbers and enclose two primes an infinite number of times. Something that holds for infinitely many numbers does not necessarily hold for all. For example, an infinite number of numbers are even, but an infinite number of numbers are not even, because they are odd. Similarly, this ruler can also be moved along the line of numbers an infinite number of times and not enclose two primes.

From Zhang's result, a deduction can be made, which is that there is a number smaller than seventy million which precisely defines a gap separating an infinite number of pairs of primes. You deduce this, a mathematician told me, by means of the pigeonhole principle. You have an infinite number of pigeons, which are pairs of primes, and you have seventy million holes. There is a hole for primes separated by two, by three, and so on. Each pigeon goes in a hole. Eventually, one hole will have an infinite number of pigeons. It isn't possible to know which one. There may even be many, there may be seventy million, but at least one hole will have an infinite number of pigeons.

Having discovered that there is a gap, Zhang wasn't interested in finding the smallest number defining the gap. This was work that he regarded as a mere technical problem, a type of manual labor—"ambulance chasing" is what a prominent mathematician called it. Nevertheless, within a week of Zhang's announcement mathematicians around the world began competing to find the lowest number. One of the observers of their activity was Terence Tao, a professor at U.C.L.A. Tao had the idea for a coöperative project in which mathematicians would work to lower the number rather than "fighting to snatch the lead," he told me.

The project, called Polymath8, started in March of 2013 and continued for about a year. Incrementally, relying also on work by a young British mathematician named James Maynard, the participants reduced the bound to two hundred and forty-six. "There are several problems with going lower," Tao said. "More and more computer power is required—someone had a high-powered computer running for two weeks to get that calculation. There were also theoretical problems. With current methods, we can never get better than six, because of something called the parity problem, which no one knows how to get past." The parity problem says that primes with certain behaviors can't be detected with current methods. "We never strongly believed we would get to two and prove the twin-prime conjecture, but it was a fun journey," Tao said.

"Is there a talent a mathematician should have?"

"Concentration," Zhang said. We were walking across the campus in a light rain. "Also, you should never give up in your personality," he continued. "Maybe something in front of you is very complicated, it's lengthy, but you should be able to pick up the major points by your intuition."

When we reached Zhang's office, I asked how he had found the door into the problem. On a whiteboard, he wrote, "Goldston-Pintz-Yıldırım"and "Bombieri-Friedlander-Iwaniec." He said, "The first paper is on bound gaps, and the second is on the distribution of primes in arithmetic progressions. I compare these two together, plus my own innovations, based on the years of reading in the library."

When I asked Peter Sarnak how Zhang had arrived at his result, he said, "What he did was look way out of reach. Maybe forty years ago the problem appeared hopeless, but in 2005 Goldston-Pintz-Yıldırım put it at the doorstep. Everybody thought, Now we're very close, but by 2011 no one was making any progress. Bombieri, Friedlander, and Iwaniec had the other important work, but it looked like you couldn't combine their ideas with Goldston. Their work was just not flexible enough to jive—it insisted on some side conditions. Then Zhang comes along. A lot of people use theorems like a computer. They think, If it is correct, then good, I'll use it. You couldn't use the Bombieri-Friedlander-Iwaniec, though, because it wasn't flexible enough. You have to take my word, because even to a serious mathematician this would be difficult to explain. Zhang understood the techniques deeply enough so as to be able to modify Bombieri-Friedlander-Iwaniec and cross this bridge. This is the most significant thing about what he has done mathematically. He's made the Bombieri-Friedlander-Iwaniec technique about the distribution of prime numbers a tool for any kind of study of primes. A development that began in the eighteen-hundreds continued through him."

"Our conditions needed to be relaxed," Iwaniec told me. "We tried, but we couldn't remove them. We didn't try long, because after failing you just start thinking there is some kind of natural barrier, so we gave up."

I asked if he was surprised by Zhang's result. "What Zhang did was sensational," he said. "His work is a masterpiece. When you talk of number theory, a lot of the beauty is the machinery. Zhang somehow completely understood the situation, even though he was working alone. That's how he surprised. He just amazingly pushed further some of the arguments in these papers."

Zhang used a very complicated form of a simple device for finding primes called a sieve, invented by a Greek named Eratosthenes, a contemporary of Archimedes. To use a simple sieve to find the primes less than a thousand, say, you write down all the numbers, then cross out the multiples of two, which can't be prime, since they are even. Then you cross out the multiples of three, then five, and so on. You have to go only as far as the multiples of thirty-one. Zhang used a different sieve from the one that others had used. The previous sieve excluded numbers once they grew too far apart. With it, Goldston, Pintz, and Yıldırım had proved that there were always two primes separated by something less than the average distance between primes that large. What they couldn't identify was a precise gap. Zhang succeeded partly by making the sieve less selective.

I asked Zhang if he was working on something new. "Maybe two or three problems I would like to solve," he said. "Bound gaps is successful, but still I have something else."

"Will it be as important?"

"Yes."

According to other mathematicians, Zhang is working on his incomplete result for the Landau-Siegel zeros conjecture. "If he succeeds, it would be much more dramatic," Peter Sarnak said. "We don't know how close he is, but he's proved that he's a genius. There's no question about that. He's also proved that he can speak with something over many years. Based on that, his chances are not zero. They're positive."

"Many people have tried that problem," Iwaniec said. "He's a private guy. Nothing is rushed. If it takes him another ten years, that's fine with him. Unless you tackle a problem that's already solved, which is boring, or one whose solution is clear from the beginning, mostly you are stuck. But Zhang is willing to be stuck much longer."

Zhang's preference for undertaking only ambitious problems is rare. The pursuit of tenure requires an academic to publish frequently, which often means refining one's work within a field, a task that Zhang has no inclination for. He does not appear to be competitive with other mathematicians, or resentful about having been simply a teacher for years while everyone else was a professor. No one who knows him thinks that he is suited to a tenure-track position. "I think what he did was brilliant," Deane Yang told me. "If you become a good calculus teacher, a school can become very dependent on you. You're cheap and reliable, and there's no reason to fire you. After you've done that a couple of years, you can do it on autopilot; you have a lot of free time to think, so long as you're willing to live modestly. There are people who try to work nontenure jobs, of course, but usually they're nuts and have very dysfunctional personalities and lives, and are unpleasant to deal with, because they feel disrespected. Clearly, Zhang never felt that."

One day, I arrived at Zhang's office as he was making tea. There was a piece of paper on his desk with equations on it and a pen on top of the paper. Zhang had an envelope in one hand. "I had a letter from an old friend," he said. "We have been separated for many years, and now he found me."

He took a pair of scissors from a drawer and cut open the envelope so slowly that he seemed to be performing a ritual. The letter was written in Chinese characters. He sat on the edge of his chair and read slowly. He put the letter down and took from the envelope a photograph of a man and a woman and a child on a sofa with a curtain in the background. He returned to reading the letter and then he put it back in the envelope and in the drawer and closed the drawer. "His new address is in Queens," he said. Then he picked up his tea and blew on it and faced me, looking at me over the top of the cup like someone peering over a wall.

I asked about Hardy's observations regarding age—Hardy also wrote, "A mathematician may still be competent enough at sixty, but it is useless to expect him to have original ideas."

"This may not apply to me," Zhang said. He put his tea on the desk and looked out the window. “Still I think I have intuition,” he said. “Still I am confident of myself. Still I have some other visions.”

Why Art Is an Asset Class

Nouriel Roubini, Roubini's Edge Feb 11, 2015

A few weeks ago, I attended the World Economic Forum in Davos, Switzerland.

Davos was crowded with the usual representatives of the top 1%: CEOs, technologists, representatives from the major banks, and the occasional head of state.

This year, there were discussions about the ECB's quantitative easing program, which was announced during the conference, and the Swiss National Bank's recent decision to remove its peg on the Swiss Franc — which sent the price of food and drinks at the hotels in Davos soaring.

I participated in twelve panel discussions and dinner engagements this year where I gave presentations. Eleven of those presentations were dedicated to the usual wonky macroeconomic policy discussions — but the twelfth was about a topic very close to my heart: The art world.

During that panel, I sketched out three main ideas:

1). That art is a new and separate asset class.
2). That there are a number of serious distortions in the art market that suggest that there is some shady behavior going on.
3). That there is a need for regulation and reform of the art market.

In fact, the second two points received a fair amount of attention in the financial press, including this excellent article by John Gapper and Peter Aspden in the Financial Times, which covers the potential for manipulation and price distortions in the art market.

What I would like to do in this newsletter is take you through the first point — why art is a unique asset class and deserves to be thought of as such.

First of all, it's important for me to say that for me art is more than just an asset class. It's impossible to reduce beauty to a simple transaction. For many years I have been a collector of work that I value for aesthetic rather than financial reasons.

With all of that said, however, there is a huge element of the art world that is related to business.

So how, specifically, do we know that art is an asset class?

1) First, the art market is extremely large. There is no way of knowing exactly how large, because there is no granular or transparent reporting of the transactions that take place there. There have been suggestions that the yearly turnover of all art sold could be as high as $70 billion, between new and existing art.

Sotheby's recently sold $423.1 million at its fall opening — its highest total ever. If you consider Sotheby's opening, and think of all the selling occurring in auction houses, galleries, estate sales, festivals, and even street markets, it is not hard to imagine the market cap of the asset class approaching an estimated $1 trillion.

(Even as I write, old records are being broken: Over the weekend, it was announced that a painting by the French Post-Impressionist artist Paul Gauguin shattered the price record for a single painting — with a sale price over $300 million.)

2) Art is used as a portfolio investment. People invest in art because they believe that there is going to be a meaningful return on their investment. Some of the investment return is driven by the preferential tax treatment that art often receives. The tax treatment varies by country but, generally speaking, art investors receive favorable capital gains taxation and benefits at the time the artwork is transferred across generations. This is especially true in the United States. The fact that the US has become the leading art market has much to do with the way we treat art with regard to capital gains and inheritance taxes.

There is, in fact, a great deal of complexity involved in the tax treatment of artwork in the US. The IRS, for example, has created an array of rules for different kinds of art owners: collectors, dealers, and investors. However, if there was any proof of art being an asset class, and a big business, perhaps the IRS stratification provides all the proof one needs.

3) The returns on art are significant. Jianping Mei and Michael Moses, the two academics behind the Mei Moses indices that track art valuations, claim that long term returns on art can be similar to the return on stock. Of course, it is tricky to calculate such comparisons with great precision. In the case of art, you have carrying costs that must be included in the calculation. (For example, insurance, conservation, transportation, and storage costs— all of which are not a factor when you invest in equities.) There are also claims that art indices suffer from selection bias, because they only capture the artists who survive and make it into the auctions. On the other hand, you have the same survivor bias in equity indices. The companies that do well make it into the Dow Jones Industrial Average and the S&P 500. Those

that don't get weeded out. So a similar type of selection bias exists with equities.

4) Art is an asset class because there are many market participants. Art galleries, auction houses, and various dealers and intermediaries take part in buying and selling art. Some of those market participants are even listed as publicly traded companies. (Sotheby's auction house, in fact, is the oldest company currently traded on the New York Stock Exchange.)

5) Art fairs have become big business. Art Basel is perhaps the best example of the growing popularity of the art world. Each year, artists, celebrities, and members of the media flock to the festival to partake in the spectacle. With tie-ins to other markets, such as the fashion industry, the popularity and commercialization of the event has only grown.

6) Private banks and financial institutions are offering their clients a variety of art related services. Those banks now have dedicated teams that are providing services to high net worth individuals and families. The services they provide include: art insurance; art storage and shipping; art conservation; and, perhaps most important, art lending.

7) Funds that invest in art are now proliferating. A recent estimate placed the number of funds actively investing in art at more than 70. Their success, to date, has been mixed, due in part to liquidity constraints in the market. One of the major challenges of investing in art is illiquidity. The absence of a truly liquid market has likely contributed to the challenges the funds face. To combat the these liquidity constraints, there has even been talk of creating a market for derivatives that track the underlying value of the assets held in these funds.

8) Traditional art auction houses are partnering with online platforms to take advantage of new opportunities in the art marketplace. The most notable example of this phenomenon is the partnership between eBay and Sotheby's, which was formed last year to expand the global reach and sales of art using digital platforms.

9) While the art market is still challenged by a meaningful amount of opacity, there has been, over time, an increasing amount of price transparency. There are, for example, public auctions that provide price discovery as well as the indices I discussed earlier. However, the market is not yet fully transparent. Regulation is still missing from the equation. In my view, however, market regulation is coming to the art world. New regulation will increase market transparency and perhaps increase turnover. Of course, art market insiders may prefer not to have this transparency. But if you view regulation from a broader perspective, rather than from the per-

spective of dealers and galleries, the impact on the art market will be, in my view, a net positive. It is more beneficial to create more trust in the asset class. Perhaps this may be in the long term interest of even the art industry insiders.

10) There is a lack of a fundamental pricing model for art. Aside from the psychic pleasure art produces, there is no clear income stream that flows from art. Other assets have both an income stream as well as a potential capital gain. Stocks provide dividends, bonds provide coupon payments, and real estate can generate rent. Art, in some ways, is closer to gold. But at the same time art is less liquid and less homogenous than gold. Copper is copper. Gold is gold. A share of Apple or GM is always the same. Meanwhile, not only is one piece of art not the same as another, but there are even different measuring standards for the same artist depending on the medium - drawings versus paintings, for example - and at what point in the artist's career the work was created.

11) This lack of a fundamental pricing model means that art is subject to fads, fashions, manias — and potentially bubbles. (Markets sometimes run into major challenges even when assets have fundamental pricing models — let alone without them.)

12) Finally, is the art market now in a bubble? Is the bubble about to burst? These questions are now being raised about the art market. As I mentioned earlier, the fact that art trades in a way that cannot be reduced to a fundamental valuation makes answering these questions more complicated. Despite this uncertainty, there are those who say, "This time is different." Those who believe that often cite the rising forces of globalization, which are now enriching a whole new class of wealthy individuals, whom they believe are going to demand art both for its aesthetic value and for investment purposes. However, it's worth mentioning that we've had booms and busts in the art world in the past. There was a major bust in the early 1990s. In the late 1980s there were a group of star artists whose stars ultimately dimmed, and the value of their art fell sharply. In the end, there are always fads. Everyone wants to own the work of the latest 'It' artist. In art, like in everything else, everyone wants the new shiny thing.

The Authenticity Paradox

Herminia Ibarra, Harvard Business Review, Feb 2015 Issue

Authenticity has become the gold standard for leadership. But a simplistic understanding of what it means can hinder your growth and limit your impact.

Consider Cynthia, a general manager in a health care organization. Her promotion into that role increased her direct reports 10-fold and expanded the range of businesses she oversaw—and she felt a little shaky about making such a big leap. A strong believer in transparent, collaborative leadership, she bared her soul to her new employees: "I want to do this job," she said, "but it's scary, and I need your help." Her candor backfired; she lost credibility with people who wanted and needed a confident leader to take charge.

Or take George, a Malaysian executive in an auto parts company where people valued a clear chain of command and made decisions by consensus. When a Dutch multinational with a matrix structure acquired the company, George found himself working with peers who saw decision making as a freewheeling contest for the best-debated ideas. That style didn't come easily to him, and it contradicted everything he had learned about humility growing up in his country. In a 360-degree debrief, his boss told him that he needed to sell his ideas and accomplishments more aggressively. George felt he had to choose between being a failure and being a fake.

Because going against our natural inclinations can make us feel like impostors, we tend to latch on to authenticity as an excuse for sticking with what's comfortable. But few jobs allow us to do that for long. That's doubly true when we advance in our careers or when demands or expectations change, as Cynthia, George, and countless other executives have discovered.

In my research on leadership transitions, I have observed that career advances require all of us to move way beyond our comfort zones. At the same time, however, they trigger a strong countervailing impulse to protect our identities: When we are unsure of ourselves or our ability to perform well or measure up in a new setting, we often retreat to familiar behaviors and styles.

But my research also demonstrates that the moments that most challenge our sense of self are the ones that can teach us the most about leading effectively. By viewing ourselves as works in progress and evolving our professional identities through trial and error, we can develop a personal style that feels right to us and suits our organizations' changing needs.

That takes courage, because learning, by definition, starts with unnatural and often superficial behaviors that can make us feel calculating instead of genuine and spontaneous. But the only way to avoid being pigeonholed and ultimately become better leaders is to do the things that a rigidly authentic sense of self would keep us from doing.

Why Leaders Struggle with Authenticity

The word "authentic" traditionally referred to any work of art that is an original, not a copy. When used to describe leadership, of course, it has other meanings—and they can be problematic. For example, the notion of adhering to one "true self" flies in the face of much research on how people evolve with experience, discovering facets of themselves they would never have unearthed through introspection alone. And being utterly transparent—disclosing every single thought and feeling—is both unrealistic and risky.

What Is Authenticity?

A too-rigid definition of authenticity can get in the way of effective leadership. Here are three examples and the problems they pose.

SOURCE HERMINIA IBARRA — HBR.ORG

Leaders today struggle with authenticity for several reasons. First, we make more-frequent and more-radical changes in the kinds of work we do. As we strive to improve our game, a clear and firm sense of self is a compass that helps us navigate choices and progress toward our goals. But when we're looking to change our game, a too rigid self-concept becomes an anchor that keeps us from sailing forth, as it did at first with Cynthia.

Second, in global business, many of us work with people who don't share our cultural norms and have different expectations for how we should behave. It can often seem as if we have to choose between what is expected—and therefore effective—

and what feels authentic. George is a case in point.

Third, identities are always on display in today's world of ubiquitous connectivity and social media. How we present ourselves—not just as executives but as people, with quirks and broader interests—has become an important aspect of leadership. Having to carefully curate a persona that's out there for all to see can clash with our private sense of self.

In dozens of interviews with talented executives facing new expectations, I have found that they most often grapple with authenticity in the following situations.

Taking charge in an unfamiliar role.

As everyone knows, the first 90 days are critical in a new leadership role. First impressions form quickly, and they matter. Depending on their personalities, leaders respond very differently to the increased visibility and performance pressure.

Psychologist Mark Snyder, of the University of Minnesota, identified two psychological profiles that inform how leaders develop their personal styles. "High self-monitors"—or chameleons, as I call them—are naturally able and willing to adapt to the demands of a situation without feeling fake. Chameleons care about managing their public image and often mask their vulnerability with bluster. They may not always get it right the first time, but they keep trying on different styles like new clothes until they find a good fit for themselves and their circumstances. Because of that flexibility, they often advance rapidly. But chameleons can run into problems when people perceive them as disingenuous or lacking a moral center—even though they're expressing their "true" chameleon nature.

By contrast, "true-to-selfers" (Snyder's "low self-monitors") tend to express what they really think and feel, even when it runs counter to situational demands. The danger with true-to-selfers like Cynthia and George is that they may stick too long with comfortable behavior that prevents them from meeting new requirements, instead of evolving their style as they gain insight and experience.

Cynthia (whom I interviewed after her story appeared in a Wall Street Journal article by Carol Hymowitz) hemmed herself in like this. She thought she was setting herself up for success by staying true to her highly personal, full-disclosure style of management. She asked her new team for support, openly acknowledging that she felt a bit at sea. As she scrambled to learn unfamiliar aspects of the business, she worked tirelessly to contribute to every decision and solve every problem. After a few months, she was on the verge of burnout. To make matters worse, sharing her vulnerability with her team members so early on had damaged her standing. Reflecting on her transition some years later, Cynthia told me: “Being authentic doesn’t mean that you can be held up to the light and people can see right through you.” But at the time, that was how she saw it—and instead of building trust, she made people question her ability to do the job.

Delegating and communicating appropriately are only part of the problem in a case like this. A deeper-seated issue is finding the right mix of distance and closeness in an unfamiliar situation. Stanford psychologist Deborah Gruenfeld describes this as managing the tension between authority and approachability. To be authoritative, you privilege your knowledge, experience, and expertise over the team’s, maintaining a measure of distance. To be approachable, you emphasize your relationships with people, their input, and their perspective, and you lead with empathy and warmth. Getting the balance right presents an acute authenticity crisis for true-to-selfers, who typically have a strong preference for behaving one way or the other. Cynthia made herself too approachable and vulnerable, and it undermined and drained her. In her bigger role, she needed more distance from her employees to gain their confidence and get the job done.

Selling your ideas (and yourself).

Leadership growth usually involves a shift from having good ideas to pitching them to diverse stakeholders. Inexperienced leaders, especially true-to-selfers, often find the process of getting buy-in distasteful because it feels artificial and political; they believe that their work should stand on its own merits.

Here’s an example: Anne, a senior manager at a transportation company, had doubled revenue and fundamentally redesigned core processes in her unit. Despite her obvious accomplishments, however, her boss didn’t consider her an inspirational leader. Anne also knew she was not communicating effectively in her role as a board member of the parent company. The chairman, a broad-brush thinker, often became impatient with her detail orientation. His feedback to her was “step up, do the vision thing.” But to Anne that seemed like valuing form over substance. “For me, it is manipulation,” she told me in an interview. “I can do the storytelling too, but I refuse to play on people’s emotions. If the string-pulling is too obvious, I can’t make myself do it.” Like many aspiring leaders, she resisted crafting emotional messages to influence and inspire others because that felt less authentic to her than relying on facts, figures, and spreadsheets. As a result, she worked at cross-purposes with the board chairman, pushing hard on the facts instead of pulling him in as a valued ally.

Many managers know deep down that their good ideas and strong potential will go unnoticed if they don't do a better job of selling themselves. Still, they can't bring themselves to do it. "I try to build a network based on professionalism and what I can deliver for the business, not who I know," one manager told me. "Maybe that's not smart from a career point of view. But I can't go against my beliefs….So I have been more limited in 'networking up.'"

Until we see career advancement as a way of extending our reach and increasing our impact in the organization—a collective win, not just a selfish pursuit—we have trouble feeling authentic when touting our strengths to influential people. True-to-selfers find it particularly hard to sell themselves to senior management when they most need to do so: when they are still unproven. Research shows, however, that this hesitancy disappears as people gain experience and become more certain of the value they bring.

Processing negative feedback.

Many successful executives encounter serious negative feedback for the first time in their careers when they take on larger roles or responsibilities. Even when the criticisms aren't exactly new, they loom larger because the stakes are higher. But leaders often convince themselves that dysfunctional aspects of their "natural" style are the inevitable price of being effective.

Let's look at Jacob, a food company production manager whose direct reports gave him low marks in a 360 review on emotional intelligence, team building, and empowering others. One team member wrote that it was hard for Jacob to accept criticism. Another remarked that after an angry outburst, he'd suddenly make a joke as if nothing had happened, not realizing the destabilizing effect of his mood changes on those around him. For someone who genuinely believed that he'd built trust among his people, all this was tough to swallow.

Once the initial shock had subsided, Jacob acknowledged that this was not the first time he'd received such criticism (some colleagues and subordinates had made similar comments a few years earlier). "I thought I'd changed my approach," he reflected, "but I haven't really changed so much since the last time." However, he quickly rationalized his behavior to his boss: "Sometimes you have to be tough in order to deliver results, and people don't like it," he said. "You have to accept that as part of the job description." Of course, he was missing the point.

Because negative feedback given to leaders often centers on style rather than skills or expertise, it can feel like a threat to their identity—as if they're being asked to give up their "secret sauce." That's how Jacob saw it. Yes, he could be explosive—

but from his point of view, his "toughness" allowed him to deliver results year after year. In reality, though, he had succeeded up to this point despite his behavior. When his role expanded and he took on greater responsibility, his intense scrutiny of subordinates became an even bigger obstacle because it took up time he should have been devoting to more-strategic pursuits.

A great public example of this phenomenon is Margaret Thatcher. Those who worked with her knew she could be merciless if someone failed to prepare as thoroughly as she did. She was capable of humiliating a staff member in public, she was a notoriously bad listener, and she believed that compromise was cowardice. As she became known to the world as the "Iron Lady," Thatcher grew more and more convinced of the rightness of her ideas and the necessity of her coercive methods. She could beat anyone into submission with the power of her rhetoric and conviction, and she only got better at it. Eventually, though, it was her undoing—she was ousted by her own cabinet.

A Playful Frame of Mind

Such a rigid self-concept can result from too much introspection. When we look only within for answers, we inadvertently reinforce old ways of seeing the world and outdated views of ourselves. Without the benefit of what I call outsight—the valuable external perspective we get from experimenting with new leadership behaviors—habitual patterns of thought and action fence us in. To begin thinking like leaders, we must first act: plunge ourselves into new projects and activities, interact with very different kinds of people, and experiment with new ways of getting things done. Especially in times of transition and uncertainty, thinking and introspection should follow experience—not vice versa. Action changes who we are and what we believe is worth doing.

Fortunately, there are ways of increasing outsight and evolving toward an "adaptively authentic" way of leading, but they require a playful frame of mind. Think of leadership development as trying on possible selves rather than working on yourself—which, let's face it, sounds like drudgery. When we adopt a playful attitude, we're more open to possibilities. It's OK to be inconsistent from one day to the next. That's not being a fake; it's how we experiment to figure out what's right for the new challenges and circumstances we face.

My research suggests three important ways to get started:

Learn from diverse role models.

Most learning necessarily involves some form of imitation—and the understanding that nothing is "original." An important part of growing as a leader is viewing authenticity not as an intrinsic state but as the ability to take elements you have learned from others' styles and behaviors and make them your own.

But don't copy just one person's leadership style; tap many diverse role models. There is a big difference between imitating someone wholesale and borrowing selectively from various people to create your own collage, which you then modify and improve. As the playwright Wilson Mizner said, copying one author is plagiarism, but copying many is research.

I observed the importance of this approach in a study of investment bankers and consultants who were advancing from analytical and project work to roles advising clients and selling new business. Though most of them felt incompetent and insecure in their new positions, the chameleons among them consciously borrowed styles and tactics from successful senior leaders—learning through emulation how to use humor to break tension in meetings, for instance, and how to shape opinion without being overbearing. Essentially, the chameleons faked it until they found what worked for them. Noticing their efforts, their managers provided coaching and mentoring and shared tacit knowledge.

As a result, the chameleons arrived much faster at an authentic but more skillful style than the true-to-selfers in the study, who continued to focus solely on demonstrating technical mastery. Often the true-to-selfers concluded that their managers were "all talk and little content" and therefore not suitable role models. In the absence of a "perfect" model they had a harder time with imitation—it felt bogus. Unfortunately, their managers perceived their inability to adapt as a lack of effort or investment and thus didn't give them as much mentoring and coaching as they gave the chameleons.

Work on getting better.

Setting goals for learning (not just for performance) helps us experiment with our identities without feeling like impostors, because we don't expect to get everything right from the start. We stop trying to protect our comfortable old selves from the threats that change can bring, and start exploring what kinds of leaders we might become.

Of course, we all want to perform well in a new situation—get the right strategy in place, execute like crazy, deliver results the organization cares about. But focusing exclusively on those things makes us afraid to take risks in the service of learning.

In a series of ingenious experiments, Stanford psychologist Carol Dweck has shown that concern about how we will appear to others inhibits learning on new or unfamiliar tasks. Performance goals motivate us to show others that we possess valued attributes, such as intelligence and social skill, and to prove to ourselves that we have them. By contrast, learning goals motivate us to develop valued attributes.

When we're in performance mode, leadership is about presenting ourselves in the most favorable light. In learning mode, we can reconcile our yearning for authenticity in how we work and lead with an equally powerful desire to grow. One leader I met was highly effective in small-group settings but struggled to convey openness to new ideas in larger meetings, where he often stuck to long-winded presentations for fear of getting derailed by others' comments. He set himself a "no PowerPoint" rule to develop a more relaxed, improvisational style. He surprised himself by how much he learned, not only about his own evolving preferences but also about the issues at hand.

Don't stick to "your story."

Most of us have personal narratives about defining moments that taught us important lessons. Consciously or not, we allow our stories, and the images of ourselves that they paint, to guide us in new situations. But the stories can become outdated as we grow, so sometimes it's necessary to alter them dramatically or even to throw them out and start from scratch.

That was true for Maria, a leader who saw herself as a "mother hen with her chicks all around." Her coach, former Ogilvy & Mather CEO Charlotte Beers, explains in I'd Rather Be in Charge that this self-image emerged from a time when Maria had to sacrifice her own goals and dreams to take care of her extended family. It eventually began to hold her back in her career: Though it had worked for her as a friendly and loyal team player and a peacekeeper, it wasn't helping her get the big leadership assignment she wanted. Together Maria and her coach looked for another defining moment to use as a touchstone—one that was more in keeping with Maria's desired future self, not who she had been in the past. They chose the time when Maria, as a young woman, had left her family to travel the world for 18 months. Acting from that bolder sense of self, she asked for—and got—a promotion that had previously been elusive.

Dan McAdams, a Northwestern psychology professor who has spent his career studying life stories, describes identity as "the internalized and evolving story that results from a person's selective appropriation of past, present and future." This isn't just academic jargon. McAdams is saying that you have to believe your story—but also embrace how it changes over time, according to what you need it to do. Try out new stories about yourself, and keep editing them, much as you would your résumé.

Again, revising one's story is both an introspective and a social process. The narratives we choose should not only sum up our experiences and aspirations but also reflect the demands we face and resonate with the audience we're trying to win over.

Countless books and advisers tell you to start your leadership journey with a clear sense of who you are. But that can be a recipe for staying stuck in the past. Your leadership identity can and should change each time you move on to bigger and better things.

The only way we grow as leaders is by stretching the limits of who we are—doing new things that make us uncomfortable but that teach us through direct experience who we want to become. Such growth doesn't require a radical personality makeover. Small changes—in the way we carry ourselves, the way we communicate, the way we interact—often make a world of difference in how effectively we lead.

Ironies of Automation

Lisanne Bainbridge, 1983

The classic aim of automation is to replace human manual control, planning and problem solving by automatic devices and computers. However, as Bibby and colleagues (1975) point out : "even highly automated systems. such as electric power networks, need human beings for supervision, adjustment, maintenance, expansion and improvement. Therefore one can draw the paradoxical conclusion that automated systems still are man-machine systems, for which both technical and human factors are important." This paper suggests that the increased interest in human factors among engineers reflects the irony that the more advanced a control system is, so the more crucial may be the contribution of the human operator.

This paper is particularly concerned with control in process industries, although examples will be drawn from flight-deck automation. In process plants the different modes of operation may be automated to different extents, for example normal operation and shut-down may be automatic while start-up and abnormal conditions are manual. The problems of the use of automatic or manual control are a function of the predictability of process behaviour, whatever the mode of operation. The first two sections of this paper discuss automatic on-line control where a human operator is expected to takeover in abnormal conditions, the last section introduces some aspects of human- computer collaboration in on-line control.

1. Introduction

The important ironies of the classic approach to automation lie in the expectations of the system designers, and in the nature of the tasks left for the human operators to carry out.

The designer's view of the human operator may be that the operator is unreliable and inefficient. so should be eliminated from the system. There are two ironies of this attitude. One is that designer errors can be a major source of operating problems. Unfortunately people who have collected data on this are reluctant to publish them, as the actual figures are difficult to interpret. (Some types of error may be reported more readily than others, and there may be disagreement about their origin.) The second irony is that the designer who tries to eliminate the operator still leaves the operator to do the tasks which the designer cannot think how to automate. It is this approach which causes the problems to be discussed here, as it means that the operator can be left with an arbitrary collection of tasks, and little thought may have been given to providing support for them.

1.1. Tasks after automation.
There are two general categories of task left for an operator in an automated system. He may be expected to monitor that the automatic system is operating correctly, and if it is not he may be expected to call a more experienced operator or to takeover himself. We will discuss the ironies of manual takeover first, as the points made also have implications for monitoring. To take over and stabilize the process requires manual control skills, to diagnose the fault as a basis for shut down or recovery requires cognitive skills.

1.1.1.Manual control skills. Several studies (Edwards and Lees, 1974) have shown the difference between inexperienced and experienced process operators making a step change. The experienced operator makes the minimum number of actions, and the process output moves smoothly and quickly to the new level, while with an inexperienced operator it oscillates round the target value. Unfortunately, physical skills deteriorate when they are not used, particularly the refinements of gain and timing. This means that a formerly experienced operator who has been monitoring an automated process may now be an inexperienced one. If he takes over he may set the process into oscillation. He may have to wait for feedback, rather than controlling by open-loop, and it will be difficult for him to interpret whether the feedback shows that there is something wrong with the system or more simply that he has misjudged his control action. He will need to make actions to counteract his ineffective control, which will add to his work load. When manual takeover is needed there is likely to be something wrong with the process, so that unusual actions will be needed to control it, and one can argue that the operator needs to be more rather than less skilled, and less rather than more loaded, than average.

1.1.2.Cognitive skills.
Long-term knowledge . An operator who finds out how to control the plant for himself, without explicit training, uses a set of propositions about possible process behaviour, from which he generates strategies to try (e.g. Bainbridge. 1981). Similarly an operator will only be able to generate successful new strategies for unusual situations if he has an adequate knowledge of the process. There are two problems with this for machine-minding operators. One is that efficient retrieval of knowledge from long-term memory depends on frequency of use (consider any subject which you passed an examination in at school and have not thought about since). The other is that this type of knowledge develops only through use and feedback about its effectiveness. People given this knowledge in theoretical classroom instruction without appropriate practical exercises will probably not understand much of it, as it will not be within a framework which makes it meaningful. and they will not remember much of it as it will not be associated with retrieval strategies which arc in-

tegrated with the rest of the task. There is some concern that the present generation of automated systems, which are monitored by former manual operators, are riding on their skills, which later generations of operators cannot be expected to have.

Working storage . The other important aspect of cognitive skills in on-line decision making is that decisions are made within the context of the operator's knowledge of the current state of the process. This is a more complex form of running memory than the notion of a limited capacity short-term store used for items such as telephone numbers. The operator has in his head (Bainbridge, 1975) not raw data about the process state, but results of making predictions and decisions about the process which will be useful in future situations, including his future actions. This information takes time to build up. Manual operators may come into the control room quarter to half an hour before they are due to take over control, so they can get this feel for what the process is doing. The implication of this for manual takeover from automatically controlled plant is that the operator who has to do something quickly can only do so on the basis of minimum information. He will not be able to make decisions based on wide knowledge of the plant state until he has had time to check and think about it.

1.1.3 Monitoring. It may seem that the operator who is expected solely to monitor that the automatics are acting correctly, and to call the supervisor if they are not, has a relatively simple task which does not raise the above complexities. One complexity which it does raise of course is that the supervisor too will not be able to takeover if he has not been reviewing his relevant knowledge, or practising a crucial manual skill. Another problem arises when one asks whether monitoring can be done by an unskilled operator.

We know from many 'vigilance' studies (Mackworth. 1950) that it is impossible for even a highly motivated human being to maintain effective visual attention towards a source of information on which very little happens, for more than about half an hour. This means that it is humanly impossible to carry out the basic function of monitoring for unlikely abnormalities, which therefore has to be done by an automatic alarm system connected to sound signals. (Manual operators will notice abnormal behaviour of variables which they look at as part of their control task, but may be equally poor at noticing changes on others.) This raises the question of who notices when the alarm system is not working properly. Again, the operator will not monitor the automatics effectively if they have been operating acceptably for a long period. A classic method of enforcing operator attention to a steady-state system is to require him to make a log. Unfortunately people can write down numbers without noticing what they are.

A more serious irony is that the automatic control system has been put in because it can do the job better than the operator, but yet the operator is being asked to monitor that it is working effectively. There are two types of problem with this. In complex modes of operation the monitor needs to know what the correct behaviour of the process should be, for example in batch processes where the variables have to follow a particular trajectory in time. Such knowledge requires either special training or special displays.

The second problem is that if the decisions can be fully specified then a computer can make them more quickly, taking into account more dimensions and using more accurately specified criteria than a human operator can. There is therefore no way in which the human operator can check in real-time that the computer is following its rules correctly. One can therefore only expect the operator to monitor the computer's decisions at some meta-level, to decide whether the computer's decisions are 'acceptable'. If the computer is being used to make the decisions because human judgement and intuitive reasoning are not adequate in this context, then which of the decisions is to be accepted? The human monitor has been given an impossible task.

1.2. Operator attitudes.
I know of one automated plant where the management had to be present during the night shift, or the operators switched the process to 'manual'. This raises general issues about the importance of skill to the individual. One result of skill is that the operator knows he can takeover adequately if required. Otherwise the job is one of the worst types, it is very boring but very responsible, yet there is no opportunity to acquire or maintain the qualities required to handle the responsibilities. The level of skill that a worker has is also a major aspect of his status, both within and outside the working community. If the job is 'deskilled' by being reduced to monitoring, this is difficult for the individuals involved to come to terms with. It also lead' to the ironies of incongruous pay differentials when the deskilled workers insist on a high pay level as the remaining symbol of status which is no longer justified by the job content

Ekkers and colleagues (19791) have published a preliminary study of the correlations between control system characteristics and the operators' subjective health and feeling of achievement. To greatly simplify : high coherence of process information, high process complexity and high process controllability (whether manual or by adequate automatics) were all associated with low levels of stress and workload and good health. and the inverse, while fast process dynamics and a high frequency of actions which cannot be made directly on the interface were associated with high stress and workload and poor health. High process controllability, good interface ergonomics and a rich pattern of activities were all associated with high feeling of achievement. Many studies show that high levels of stress lead to errors, while poor health and low job satisfaction lead to the high indirect costs of absenteeism, etc. (e.g. Mobley and colleagues, 1979).

2. Approaches to solutions

One might state these problems as a paradox that by automating the process the human operator is given a task which is only possible for someone who is in on-line control. This section will discuss some possible solutions to problems of maintaining the efficiency and skills of the operator if he is expected to monitor and take over control : the next section will introduce recent proposals for keeping the human operator on line with computer support.

Solving these problems involves very multidimensional decision making: suggestions for discussion will be made here. The recommendations in any particular case will depend on such factors as process size and complexity, the rate of process change, the speed and frequency of process or automatic control failure, the variability of the product and the environment, the simplicity and cost of shut down, and the qualities of the operator.

2.1. Monitoring.

In any situation where a low probability event must be noticed quickly then the operator must be given artificial assistance, if necessary even alarms on alarms. In a process with a large number of loops there is no way in which the human operator can get quickly to the correct part of the plant without alarms, preferably also some form of alarm analysis. Unfortunately a proliferation of flashing red lights will confuse rather than help. There are major problems and ironies in the design of large alarm systems for the human operator (Rasmussen and Rouse, 1981).

Displays can help the operator to monitor automatic control performance, by showing the target values. This is simple for single tolerance bands, but becomes more complex if tolerances change throughout batch processing. One possible solution is to show the currently appropriate tolerances on a VDU by software generation. This does not actually get round the problems, but only raises the same ones in a different form. The operator will not watch the VDU if there is a very low probability of the computer control failing. If the computer can generate the required values then it should also be able to do the monitoring and alarms. And how does the operator monitor that the computer is working correctly, or take over if it obviously is not? Major problems may be raised for an operator who is highly practised at using computer generated displays if these are no longer available in an emergency. One ironic but sensible suggestion is that direct wired displays should be used for the main process information, and software displays for quantitative detail (Jervis and Pope, 1977).

'Catastrophic' breaks to failure are relatively easy to identify. Unfortunately automatic control can 'camouflage' system failure by controlling against the variable changes, so that trends do not become apparent until they are beyond control. This implies that the automatics should also monitor unusual variable movement. 'Graceful degradation' of performance is quoted in 'Fitts Lists' of man-computer qualities as an advantage of man over machine. This is not an aspect of human performance to be aimed for in computers, as it can raise problems with monitoring for failure (e.g. Wiener and Curry. 1980), automatic systems should fail obviously.

If the human operator must monitor the details of computer decision making then, ironically, it is necessary for the computer to make these decisions using methods and criteria, and at a rate, which the operator can follow, even when this may not be the most efficient method technically. If this is not done then when the operator does not believe or agree with the computer he will be unable to trace back through the system's decision sequence to see how far he does agree.

One method of overcoming vigilance problems which is frequently suggested is to increase the signal rate artificially. It would be a mistake, however, to increase artificially the rate of computer failure as the operator will then not trust the system. Ephrath (1980) has reported a study in which system performance was worse with computer aiding, because the operator made the decisions anyway, and checking the computer added to his workload.

2.2. Working storage.

If the human operator is not involved in on-line control he will not have detailed knowledge of the current state of the system. One can ask what limitations this places on the possibility for effective manual takeover, whether for stabilization or shutdown of the process, or for fault diagnosis.

The straightforward solution when shutdown is simple and low-cost is to shut down automatically. The problems arise with processes which, because of complexity, cost or other factors (e.g. an aircraft in the air) must be stabilized rather than shutdown. Should this be done manually or automatically? Manual shut down is usable if the process dynamics can be left for several minutes while the operator works out what is happening. For very fast failures, within a few seconds (e.g. pressurized water nuclear reactor rather than an aircraft), when there is no warning from prior changes so that on-line working storage would also be useless, then reliable automatic response is necessary, whatever the investment needed, and if this is not possible then the process should not be built if the costs of failure are unacceptable.

With less fast failures it may be possible to ‘buy time’ with overlearned manual responses. This requires frequent practice on a high fidelity simulator, and a sufficient understanding of system failures to be sure that all categories of failure are covered. If response to failure requires a larger number of separate actions than can be made in the time available then some must be made automatically and the remainder by a highly practised operator.

2.3. Long-term knowledge.
Points in the previous section make it clear that it can be important to maintain manual skills. One possibility is to allow the operator to use hands-on control for a short period in each shift. If this suggestion is laughable then simulator practice must be provided. A simulator adequate to teach the basic behaviour of the process can be very primitive. Accurate fast reactions can only be learned on a high fidelity simulator, so if such reactions are necessary then this is a necessary cost.

Similar points can be made about the cognitive skills of scheduling and diagnosis. Simple pictorial representations are adequate for training some types of fault detection (Duncan and Shepherd, 1975), but only if faults can be identified from the steady-state appearance of the control panel. and waiting for the steady-state is acceptable. If fault detection involves identifying changes over time then dynamic simulators are needed for training (Marshall and Shepherd, 1981). Simple recognition training is also not sufficient to develop skills for dealing with unknown faults or for choosing corrective actions (Duncan, 1981).

There are problems with the use of any simulator to train for extreme situations. Unknown faults cannot be simulated, and system behaviour may not be known for faults which can be predicted but have not been experienced. This means that training must be concerned with general strategies rather than specific responses; for example simulations can be used to give experience with low probability events, which may be known to the trainer but not to the trainee. No one can be taught about unknown properties of the system, but they can be taught to practise solving problems within the known information. It is inadequate to expect the operator to react to unfamiliar events solely by consulting operating procedures. These cannot cover all the possibilities, so the operator is expected to monitor them and fill in the gaps. However, it is ironic to train operators in following instructions and then put them in the system to provide intelligence.

Of course, if there are frequent alarms throughout the day then the operator will have a large amount of experience of controlling and thinking about the process as part of his normal work. Perhaps the final irony is that it is the most successful automated systems, with rare need for manual intervention, which may need the greatest investment in human operator training.

3. Human-computer collaboration

By taking away the easy parts of his task. automation can make the difficult parts of the human operator's task more difficult. Several writers (Wiener and Curry, 1980; Rouse. 1981) point out that the 'Fitts list' approach to automation. assigning to man and machine the tasks they are best at, is no longer sufficient. It does not consider the integration of man and computer, nor how to maintain the effectiveness of the human operator by supporting his skills and motivation. There will always be a substantial human involvement with automated systems because criteria other than efficiency, are involved, e.g. when the cost of automating some modes of operation is not justified by the value of the product, or because the public will not accept high-risk systems with no human component. This suggests that methods of human computer collaboration need to be more fully developed. DelIner (1981) lists the possible levels of human intervention in automated decision making. This paper will discuss the possibilities for computer intervention in human decision making. These include instructing or advising the operator, mitigating his errors, providing sophisticated displays, and assisting him when task loads are high. Rouse (1981) calls these 'covert' human-computer interaction.

3.1. Instructions and advice.

Using the computer to give instructions is inappropriate if the operator is simply acting as a transducer, as the computer could equally well activate a more reliable one. Thompson (1981) lists four types of advice, about : underlying causes, relative importance, alternative actions available, and how to implement actions. When following advice the operator's reactions will be slower, and less integrated than if he can generate the sequence of activity himself, and he is getting no practice in being 'intelligent'. There are also problems with the efficient display of procedural information.

3.2. Mitigating human error.

Machine possibilities for counteracting human error range from simple hardware interlocks to complex on-line computation. Except where specific sequences of operations must be followed it is more appropriate to place such 'checks' on the effects of actions, as this does not make assumptions about the strategy used to reach this effect. Under manual control human operators often obtain enough feedback about the results of their actions within a few seconds to correct their own errors (Ruffell-Smith. 1979), but Wiener and Curry (1980) give examples of humans making the same types of errors in setting up and monitoring automatic equipment, when they do not get adequate feedback. This should perhaps be designed in. Kreifeldt and McCarthy (1981) give advice about displays to help operators who have been interrupted in mid-sequence. Rouse (1981) suggests computer monitoring of human eye movements to check that instrument scanning is appropriate, for example to prevent tunnel vision.

3.3. Software generated displays .

The increasing availability of soft displays on VDUs raises fascinating possibilities for designing displays compatible with the specific knowledge and cognitive processes being used in a task. This has led to such rich veins of creative speculation that it seems rather mean to point out that there are difficulties in practice.

One possibility is to display only data relevant to a particular mode of operation. such as start up routine operations, or maintenance. Care is needed however, as it is possible for an interface which is ideal for normal conditions to camouflage the development of abnormal ones (Edwards. 1981).

Goodstein (1981) has discussed process displays which are compatible with different types of operator skill, using a classification of three levels of behaviour suggested by Rasmussen (1979), i.e. skill based, rule based and knowledge based. The use of different types of skill is partly a function of the operator's experience, though the types probably do not fall on a simple continuum. Chafin (198l) has discussed how interface design recommendations depend on whether the operator is naive, novice/competent, or expert. However, he was concerned with human access to computer data bases when not under time pressure. Man-machine interaction under time pressure raises special problems. The change between knowledge-based thinking and reflex reaction is not solely a function of practice, but also depends on the uncertainty of the environment, so that the same task elements may be done using different types of skill at different times. It could therefore confuse rather than help the operator to give him a display which is solely a function of his overall skill level. Non-time-stressed operators, if they find they have the wrong type of display, might themselves request a different level of information. This would add to the work load of someone making decisions which are paced by a dynamic system. Rouse (1981) has therefore suggested that the computer might identify which type of skill the operator is using, and change the displays (he does not say how this might be done). We do not know how confused operators would be by display changes which were not under their own control. Ephraph and Young (1981) have commented that it takes time for an operator to shift between activity modes, e.g. from monitoring to controlling, even when these are under his control, and one assumes that the same problems would arise with changes in display mode. Certainly a great deal of care would be needed to make sure that the different displays were compatible. Rasmussen and Lind's recent paper (1981) was about the different levels of abstraction at which the operator might be thinking about the process, which would define the knowledge base to be displayed. Again, although operators evidently do think at different levels of complexity and abstraction at different times, it is not clear that they would be able to use, or choose, many different displays under time stress.

Some points were made above about the problems of operators who have learned to work with computer generated displays, when these displays are no longer available in abnormal conditions. Recent research on human memory (Craik, 1979) suggests that the more processing for meaning that some data has received the more effectively it is remembered. This makes one wonder how much the operator will learn about the structure of the process if information about it is presented so successfully that he does not have to think about it to take it in. It certainly would be ironic if we find that the most compatible display is not the best display to give to the operator after all! (As usual with display choice decisions this would depend on the task to be done. A highly compatible display always supports rapid reactions. These points speculate whether they also support acquisition of the knowledge and thinking skills needed in abnormal conditions.)

A few practical points can be suggested. There should be at least one source of information permanently available for each type of information which cannot be mapped simply onto others, e.g. about layout of plant in space as opposed to its functional topology. Operators should not have to page between displays to obtain information about abnormal states in parts of the process other than the one they are currently thinking about, nor between displays giving information needed within one decision process. Research on sophisticated displays should concentrate on the problems of ensuring compatibility between them, rather than finding which independent display is best for one particular function without considering its relation to information for other functions. To end on a more optimistic note, software displays offer some interesting possibilities for enriching the operator's task by allowing him to design his own interface.

3.4. Relieving human workload.

A computer can be used to reduce human workload either by simplifying the operator's decisions. as above, or by taking over some of the decision making. The studies which have been done on this show that it is a complex issue. Ephrath and Young (1981) found that overall control performance was better with manual control of a single loop, but was also better with an autopilot in the complex environment of a cockpit simulator. This suggests that aiding is best used at higher work loads. However. the effect of the type of aiding depends on the type of workload. Johannsen and Rouse (1981) found that pilots reported less depth of planning under autopilot in abnormal environmental conditions, presumably because the autopilot was dealing with the conditions, but more planning under emergency aircraft conditions, where they suggest that the autopilot frees the pilot from on-line control so he can think about other things. Chu and Rouse (19791 studied a situation with both computer aiding and autopilot. They arranged for the computer to take over decision making when the operator had a queue of one other task item to be dealt with and he was controlling manually, or after a queue of three items if the autopilot was controlling. The study by Enstrom and Rouse (1977) makes it clear why Rouse (l98l) comments that more sophisticated on-line methods of adapting computer aiding to human workload will only be possible if the workload computations can be done in real time. It would be rash to claim it as an irony that the aim of aiding human limited capacity has pushed computing to the limit of its capacity. as technology has a way of catching up with such remarks. Enstrom and Rouse also make the important point that the human being must know which tasks the computer is dealing with and how. Otherwise the same problems arise as in human teams in which there is no clear allocation of responsibility. Sinaiko (1972) makes a comment which emphasizes the importance of the human operator's perception of the computer's abilities: "when loads were light. the man appeared willing to let the computer carry most of the assignment responsibility: when loads were heavy, the men much more often stepped in and over-rode the computer". Evidently, quite apart from technical considerations, the design of computer aiding is a multidimensional problem.

4. Conclusion

The ingenious suggestions reviewed in the last section show that humans working without time-pressure can be impressive problem solvers. The difficulty remains that they are less effective when under time pressure. I hope this paper has made clear both the irony that one is not by automating necessarily removing the difficulties, and also the possibility that resolving them will require even greater technological ingenuity than does classic automation.