

The Computer and the Hernia Factory

One summer day in 1996, Hans Ohlin, the fifty-year-old chief of coronary care at the University of Lund Hospital in Sweden, sat down in his office with a stack of two thousand two hundred and forty electrocardiograms. Each test result consisted of a series of wavy lines, running from left to right on a letter-size page of graph paper. Ohlin read them alone in his office so that he would not be disturbed. He scanned them swiftly but carefully, one at a time, separating them into two piles according to whether or not he thought that the patient was having a heart attack at the time the electrocardiogram (EKG) was recorded. To avoid fatigue and inattention, he did his work over the course of a week, sorting through the EKGs in shifts no longer than two hours, and taking long breaks. He wanted no careless errors; the stakes were too high. This was the medical world's version of the Deep Blue chess match, and Ohlin was cardiology's Gary Kasparov. He was going head to head with a computer.

The EKG is one of the most common of diagnostic tests, performed more than fifty million times a year in the United States alone. Electrodes are placed on the skin to pick up the low-voltage electrical impulses that, with each beat, travel through the heart muscle, and those impulses are reflected in the waves on an EKG

printout. The theory behind an EKG is that in a heart attack a portion of the muscle dies, causing the electrical impulses to change course when they travel around the dead tissue. As a result, the waves on the printout change, too. Sometimes those changes are obvious; more often they are subtle—or, in medical argot, “nonspecific.”

To medical students, EKGs seem unmanageably complex at first. Typically, an EKG uses twelve leads, and each one produces a different-looking tracing on the printout. Yet students are taught to discern in these tracings a dozen or more features, each of which is given an alphabetical label: for instance, there's the downstroke at the start of a beat (the Q wave), the upstroke at the peak of heart contraction (the R wave), the subsequent downstroke (the S wave), and the rounded wave right after the beat (the T wave). Sometimes small changes here and there add up to a heart attack; sometimes they don't. When I was a medical student, I first learned to decode the EKG as if it were a complex calculation. My classmates and I would carry laminated cards in our white-lab-coat pockets with a list of arcane instructions: calculate the heart rate and the axis of electrical flow, check for a rhythm disturbance, then check for an ST-segment elevation greater than one millimeter in leads V₁ to V₄, or for poor R-wave progression (signifying one type of heart attack), and so on.

With practice, it gets easier to manage all this information, just as putting a line in gets easier. The learning curve operates in matters of diagnosis no less than technique. An experienced cardiologist can sometimes make out a heart attack at a glance, the way a child can recognize his mother across a room. But at bottom the test remains stubbornly opaque. Studies have shown that between 2 and 8 percent of patients with heart attacks who are seen in emergency rooms are mistakenly discharged, and a quarter of these people die or suffer a complete cardiac arrest. Even if such patients aren't mistakenly sent home, crucial treatment may be delayed when an EKG is misread. Human judgment, even expert human judgment, falls well short of certainty. The rationale for trying to teach a computer to read an EKG, therefore, is fairly compelling. If the result should prove to be

even a slight improvement on human performance, thousands of lives could be saved each year.

The first suggestion that a computer could do better came in 1990, in an influential article published by William Baxt, then an emergency physician at the University of California at San Diego. Baxt described how an “artificial neural network”—a kind of computer architecture—could make sophisticated clinical decisions. Such expert systems learn from experience much as humans do: by incorporating feedback from each success and each failure to improve their guesswork. In a later study, Baxt showed that a computer could handily outperform a group of doctors in diagnosing heart attacks among patients with chest pain. But two-thirds of the physicians in his study were inexperienced residents, whom you'd expect to have difficulties with EKGs. Could a computer outperform an experienced specialist?

This question was what the Swedish study was trying to answer. The study was led by Lars Edenbrandt, a medical colleague of Ohlin's and an expert in artificial intelligence. Edenbrandt spent five years perfecting his system, first in Scotland and then in Sweden. He fed his computer EKGs from more than ten thousand patients, telling it which ones represented heart attacks and which ones did not, until the machine grew expert at reading even the most equivocal of EKGs. Then he approached Ohlin, one of the top cardiologists in Sweden and a man who ordinarily read as many as ten thousand EKGs a year. Edenbrandt selected two thousand two hundred and forty EKGs from the hospital files to test both of them on, of which exactly half, eleven hundred and twenty, were confirmed to show heart attacks. With little fanfare, the results were published in the fall of 1997. Ohlin correctly picked up six hundred and twenty. The computer picked up seven hundred and thirty-eight. Machine beat man by 20 percent.

Western medicine is dominated by a single imperative—the quest for machine-like perfection in the delivery of care. From the first day of medical training, it is clear that errors are unacceptable.

Taking time to bond with patients is fine, but every X ray must be tracked down and every drug dose must be exactly right. No allergy or previous medical problem can be forgotten, no diagnosis missed. In the operating room, no movement, no time, no drop of blood can be wasted.

The keys to this kind of perfection are routinization and repetition: survival rates after heart surgery, vascular surgery, and other operations are directly related to the number of procedures the surgeon has performed. Twenty-five years ago, general surgeons performed hysterectomies, removed lung cancers, and bypassed hardened leg arteries. Today, each condition has its specialists, who perform one narrow set of procedures over and over again. When I'm in the operating room, the highest praise I can receive from my fellow surgeons is "You're a machine, Cawande." And the use of "machine" is more than casual: human beings, under some circumstances, really can act like machines.

Consider a relatively simple surgical procedure, a hernia repair, which I learned to do as a first-year surgical resident. A hernia is a weakening of the abdominal wall, usually in the groin, that allows the abdomen's contents to bulge through. In most hospitals, fixing it—pushing the bulge back in and repairing the abdominal wall—takes about ninety minutes and might cost upward of four thousand dollars. In anywhere from 10 to 15 percent of the cases, the operation eventually fails and the hernia returns. There is, however, a small medical center outside Toronto, known as the Shouldice Hospital, where none of these statistics apply. At Shouldice, hernia operations often take from thirty to forty-five minutes. Their recurrence rate is an astonishing 1 percent. And the cost of an operation is about half of what it is elsewhere. There's probably no better place in the world to get a hernia repaired.

What's the secret of that clinic's success? The short answer is that the dozen surgeons at Shouldice do hernia operations and nothing else. Each surgeon repairs between six hundred and eight hundred hernias a year—more than most general surgeons do in a lifetime. In

this particular field, Shouldice's staff is better trained and has more experience than anyone else. But there's another way to formulate the reason for its success, which is that all the repetition changes the way they think. As Lucian Leape, a Harvard pediatric surgeon who has made a study of medical error, explains, "a defining trait of experts is that they move more and more problem-solving into an automatic mode." With repetition, a lot of mental functioning becomes automatic and effortless, as when you drive a car to work. Novel situations, however, usually require conscious thought and "workaround" solutions, which are slower to develop, more difficult to execute, and more prone to error. A surgeon for whom most situations have automatic solutions has a significant advantage. If the Swedish EKG study argues that there are situations in which machines should replace physicians, the Shouldice example suggests that physicians should be trained to act more like machines.

One chilly Monday morning, I put on a green cotton scrub top and pants, a disposable mask, and a paper cap, and wandered among cases in the Shouldice Hospital's five operating rooms. To describe one case is to describe them all: I watched three surgeons operate on six patients, and none deviated even a step from their standard protocol. In a tiled, boxlike operating room, I peered over the shoulder of Richard Sang, a fifty-one-year-old surgeon with a dry wit and a youthful appearance. Though we chatted during the entire operation, Dr. Sang performed each step without pause, almost absently, with the assistant knowing precisely which tissues to retract, and the nurse handing over exactly the right instruments; instructions were completely unnecessary. The patient, a pleasant, surprisingly composed man of about thirty-five, who occasionally piped up from under the drapes to ask how things were going, lay on the table with his lower abdomen exposed and painted yellow with a bactericidal iodine solution. A plum-size bulge was visible to the left side of the hard bone of the pubis. Dr. Sang injected the skin with a local anesthetic in a diagonal line from the top of the man's left hip to the pubis, along the crease of the groin. With a No. 10 blade, he made a

four-inch slash along this line in a single downstroke, revealing yellow, glistening fat below. The assistant laid a cloth along each side of the wound to absorb the mild bleeding, and pulled it open.

Sang swiftly cut down through the outer muscle layer of the abdominal wall, exposing the spermatic cord, a half-inch cable of blood and spermatic vessels. The patient's bulge, we could now see, came through a weakness in the muscle wall beneath the cord, which is a common site. Sang slowed down for a moment, checking meticulously for another hernia, along the area where the cord came through the inner abdominal wall. Sure enough, he found a small, second hernia there—one that, if it had been missed, would almost certainly have caused a recurrence. He then sliced open the remaining muscle layers beneath the cord, so that the abdominal wall was completely open, and pushed the bulging abdominal contents back inside. If you have a tear in a couch cushion with stuffing coming through it, you can put a patch on the cushion or you can sew it back together. At my hospital, we usually push the hernia back in, place a piece of sturdy plasticlike mesh on top, and sew it to the surrounding tissue. It provides a reliable reinforcement, and the technique is easy to perform. But Sang, like the other Shouldice surgeons I asked, scoffed at the idea: they viewed the mesh as a hazard for infection (since it's a foreign body), expensive (since the mesh can cost hundreds of dollars), and unnecessary (since they get enviable results without it).

As Sang and I talked about such alternatives, he sewed the wall back together in three separate muscle layers, using fine wire, making sure that the edge of each layer overlapped like a double-breasted suit. After Sang closed the patient's skin with small clips and removed the drapes, the patient swung his legs over the edge of the table, stood up, and walked out of the room. The procedure had taken just half an hour.

Many surgeons elsewhere use Shouldice's distinctive repair method but obtain ordinary rates of recurrence. It's not the tech-

nique that makes Shouldice great. The doctors at Shouldice deliver hernia repairs the way Intel makes chips: they like to call themselves a "focused factory." Even the hospital building is specially designed for hernia patients. Their rooms have no phones or televisions, and their meals are served in a downstairs dining hall; as a result, the patients have no choice but to get up and walk around, thereby preventing problems associated with inactivity, such as pneumonia or leg clots.

After Sang left the patient with a nurse, he found the next patient and walked him straight back into the same operating room. Hardly three minutes had passed, but the room was already clean. Fresh sheets and new instruments were already laid out. And so the next case began. I asked Byrnes Shouldice, a son of the clinic's founder and a hernia surgeon himself, whether he ever got bored doing hernias all day long. "No," he said in a Spock-like voice. "Perfection is the excitement."

Paradoxically, this kind of superspecialization raises the question of whether the best medical care requires fully trained doctors. None of the three surgeons I watched operate at the Shouldice Hospital would even have been in a position to conduct their own procedures in a typical American hospital, for none had completed general surgery training. Sang was a former family physician; Byrnes Shouldice had come straight from medical school; and the surgeon-in-chief was an obstetrician. Yet after apprenticing for a year or so they were the best hernia surgeons in the world. If you're going to do nothing but fix hernias or perform colonoscopies, do you really need the complete specialists' training (four years of medical school, five or more years of residency) in order to excel? Depending on the area of specialization, do you—and this is the question posed by the Swedish EKG study—even have to be human?

Although the medical establishment has begun to recognize that automation like the Shouldice's may be able to produce better

results in medical treatment, many doctors are not fully convinced. And they have been particularly reluctant to apply the same insight to the area of medical diagnosis. Most physicians believe that diagnosis can't be reduced to a set of generalizations—to a "cookbook," as some say. Instead, they argue, it must take account of the idiosyncrasies of individual patients.

This only stands to reason, doesn't it? When I am the surgical consultant in the emergency department, I'm often asked to assess whether a patient with abdominal pain has appendicitis. I listen closely to his story and consider a multitude of factors: how his abdomen feels to me, the pain's quality and location, his temperature, his appetite, the laboratory results. But I don't plug it all into a formula and calculate the result. I use my clinical judgment—my intuition—to decide whether he should undergo surgery, be kept in the hospital for observation, or be sent home. We've all heard about individuals who defy the statistics—the hardened criminal who goes straight, the terminal cancer patient who miraculously recovers. In psychology, there's something called the broken-leg problem. A statistical formula may be highly successful in predicting whether or not a person will go to a movie in the next week. But someone who knows that this person is laid up with a broken leg will beat the formula. No formula can take into account the infinite range of such exceptional events. That's why doctors are convinced that they'd better stick with their well-honed instincts when they're making a diagnosis.

One weekend on duty, I saw a thirty-nine-year-old woman with pain in the right-lower abdomen who did not fit the pattern for appendicitis. She said that she was fairly comfortable and she had no fever or nausea. Indeed, she was hungry, and she did not jump when I pressed on her abdomen. Her test results were largely equivocal. But I still recommended appendectomy to the attending surgeon. Her white blood cell count was high, suggesting infection, and, moreover, she just looked sick to me. Sick patients can have a cer-

tain unmistakable appearance you come to recognize after a while in residency. You may not know exactly what is going on, but you're sure it's something worrisome. The attending physician accepted my diagnosis, operated, and found appendicitis.

Not long after, I had a sixty-five-year-old patient with almost precisely the same story. The lab findings were the same; I also got an abdominal scan, but it was inconclusive. Here, too, the patient didn't fit the pattern for appendicitis; here, too, he just looked to me as if he had it. In surgery, however, the appendix turned out to be normal. He had diverticulitis, a colon infection that usually doesn't require an operation.

Is the second case more typical than the first? How often does my intuition lead me astray? The radical implication of the Swedish study is that the individualized, intuitive approach that lies at the center of modern medicine is flawed—it causes more mistakes than it prevents. There's ample support for this conclusion from studies outside medicine. Over the past four decades, cognitive psychologists have shown repeatedly that a blind algorithmic approach usually trumps human judgment in making predictions and diagnoses. The psychologist Paul Meehl, in his classic 1954 treatise, *Clinical Versus Statistical Prediction*, described a study of Illinois parolees that compared estimates given by prison psychiatrists that a convict would violate parole with estimates derived from a rudimentary formula that weighed such factors as age, number of previous offenses, and type of crime. Despite the formula's crudeness, it predicted the occurrence of parole violations far more accurately than the psychiatrists did. In recent articles, Meehl and the social scientists David Faust and Robyn Dawes have reviewed more than a hundred studies comparing computers or statistical formulas with human judgment in predicting everything from the likelihood that a company will go bankrupt to the life expectancy of liver-disease patients. In virtually all cases, statistical thinking equaled or surpassed human judgment. You might think that a human being and a computer working

together would make the best decisions. But, as the researchers point out, this claim makes little sense. If opinions agree, no matter. If they disagree, the studies show that you're better off sticking with the computer's judgment.

What accounts for the superiority of a well-developed computer algorithm? First, Dawes notes, human beings are inconsistent: we are easily influenced by suggestion, the order in which we see things, recent experience, distractions, and the way information is framed. Second, human beings are not good at considering multiple factors. We tend to give some variables too much weight and wrongly ignore others. A good computer program consistently and automatically gives each factor its appropriate weight. After all, Meehl asks, when we go to the store, do we let the clerk eyeball our groceries and say, "Well, it looks like seventeen dollars' worth to me"? With lots of training, the clerk might get very good at guessing. But we recognize the fact that a computer that simply adds up the prices will be more consistent and more accurate. In the Swedish study, as it turned out, Ohlin rarely made obvious mistakes. But many EKGs are in the gray zone, with some features suggesting a healthy heart and others suggesting a heart attack. Doctors have difficulty estimating faithfully which way the mass of information tips, and they are easily influenced by extraneous factors, such as what the last EKG they came across looked like.

It is probably inevitable that physicians will have to let computers take over at least some diagnostic decisions. One network, PAPER, has already gained mainstream use in the screening of digitized Pap smears—microscopic scrapings taken from a woman's cervix—for cancer or precancerous abnormalities, which is a job usually done by a pathologist. Researchers have completed more than a thousand studies on the use of neural networks in nearly every field of medicine. Networks have been developed to diagnose appendicitis, dementia, psychiatric emergencies, and sexually transmitted diseases. Others can predict success from cancer treatment, organ transplantation, and heart valve surgery. Systems have been designed

to read chest X rays, mammograms, and nuclear-medicine heart scans.

In the treatment of disease, parts of the medical world have already begun to extend the lesson of the Shouldice Hospital concerning the advantages of specialized, automated care. Regina Herzinger, a professor at the Harvard Business School, who introduced the term "health-care focused factory" in her book *Market-Driven Health Care*, points to other examples, including the Texas Heart Institute for cardiac surgery and Duke University's bone-marrow transplant center. Breast cancer patients seem to do best in specialized cancer treatment centers, where they have a cancer surgeon, an oncologist, a radiation therapist, a plastic surgeon, a social worker, a nutritionist, and others who see breast cancer day in and day out. And almost any hospital one goes to now has protocols and algorithms for treating at least a few common conditions, such as asthma or sudden stroke. The new artificial neural networks merely extend these lessons to the realm of diagnosis.

Still, resistance to this vision of mechanized medicine will remain. Part of it may well be short-sightedness: doctors can be stubborn about changing the way we do things. Part of it, however, stems from legitimate concern that, for all the technical virtuosity gained, something vital is lost in medicine by machine. Modern care already lacks the human touch, and its technocratic ethos has alienated many of the people it seeks to serve. Patients feel like a number too often as it is.

Yet compassion and technology aren't necessarily incompatible; they can be mutually reinforcing. Which is to say that the machine, oddly enough, may be medicine's best friend. On the simplest level, nothing comes between patient and doctor like a mistake. And while errors will always dog us—even machines are not perfect—trust can only increase when mistakes are reduced. Moreover, as "systems" take on more and more of the technical work of medicine, individual physicians may be in a position to embrace the dimensions of care that mattered long before technology came—like talking to their

patients. Medical care is about our life and death, and we've always needed doctors to help us understand what is happening and why, and what is possible and what is not. In the increasingly tangled web of experts and expert systems, a doctor has an even greater obligation to serve as a knowledgeable guide and confidant. Maybe machines can decide, but we still need doctors to heal.