The unreasonable effectiveness of my self-experimentation

Seth Roberts *

Tsinghua University, Beijing, China
University of California, Berkeley, CA, United States

ARTICLE INFO

SUMMARY

Over 12 years, my self-experimentation found new and useful ways to improve sleep, mood, health, and weight. Why did it work so well? First, my position was unusual. I had the subject-matter knowledge of an insider, the freedom of an outsider, and the motivation of a person with the problem. I did not need to publish regularly. I did not want to display status via my research. Second, I used a powerful tool. Self-experimentation about the brain can test ideas much more easily (by a factor of about 500,000) than conventional research about other parts of the body. When you gather data, you sample from a power-law-like distribution of progress. Most data helps a little; a tiny fraction of data helps a lot. My subject-matter knowledge and methodological skills (e.g., in data analysis) improved the distribution from which I sampled (i.e., increased the average amount of progress per sample). Self-experimentation allowed me to sample from it much more often than conventional research. Another reason my self-experimentation was unusually effective is that, unlike professional science, it resembled the exploration of our ancestors, including foragers, hobbyists, and artisans.

© 2010 Elsevier Ltd. All rights reserved.

Introduction

In 1960, a physicist named Eugene Wigner published an essay called "The Unreasonable Effectiveness of Mathematics in the Natural Sciences" [1]. Mathematics invented to describe one thing, said Wigner, had often turned out to provide a good description of something much different. He could not explain this. I have been puzzled in a similar way. Over 12 years (1990–2002), my self-experimentation found new ways to improve sleep, mood, health, and weight [2]. Four of the new ways (avoiding breakfast to reduce early awakening, seeing morning faces to improve mood, standing to reduce early awakening, and drinking sugar water to lose weight) were surprising; almost all were practical. In health, as in other areas of science, we expect progress to come from subject-matter experts (e.g., sleep researchers) with grants. Yet I was not an expert in what I studied and my research cost almost nothing. I did it in my spare time. In spite of this, my self-experimental research was far better than my mainstream research (e.g., [3,4]). For a long time, this puzzled me. Now I propose an explanation.

The puzzle began in graduate school, where I studied experimental psychology. To learn how to do experiments, I tried to do as many as possible. At the time I had acne. It was easy to measure (count pimples each morning) so I decided to do experiments about it. My dermatologist had prescribed tetracycline, an antibiotic. A few months of self-experimentation showed that tetracy-

cline did not work [5], which surprised my dermatologist. Later conventional research found that tetracycline often fails [6,7]. My dermatologist had years of experience. Yet a little self-experimentation by a non-expert found something important that he and other dermatologists did not know.

I continued to self-experiment. I made little progress the first 10 years but plenty of progress the next 12 [2]. The new ideas suggested by my results included (from more to less important): 1. A theory about mood and depression. 2. A theory about weight control. 3. Seeing faces in the morning makes mood worse that evening and better the next day. 4. Sugar water causes weight loss. 5. Better sleep can greatly reduce colds. 6. Breakfast can cause early awakening. 7. Standing a lot reduces early awakening. 8. Ways to change the size of the faces effect. 9. Other weight-loss methods. Most of my conclusions were also supported by conventional research. Some of the weight-loss methods (such as drinking lots of water or eating lots of sushi) could not be sustained, but the rest were useful.

No well-known idea fully explains how much progress I made. Better equipment (e.g., microscope, interferometer) often produces new scientific ideas. I had used a personal computer (new at the time), which made data collection much easier. However, this was not sufficient to explain my unusual progress because most other scientists had also started to use personal computers. Kuhn [8] argued that the accumulation of inexplicable results helps generate progress. There is some truth to this. My weight-control theory was partly inspired by the hard-to-explain results of Ramirez [9]. I thought of my theory soon after learning about them. It was not that simple, however; I also did four experiments that

* Address: 1152 Arch Street, Berkeley, CA 94708, United States. Tel.: +1 510 418 7753.
E-mail address: twoutopias@gmail.com

supported the new theory. Moreover, my mood theory did not fit Kuhn's pattern. It had nothing to do with inexplicable results. It was an obvious conclusion from the effects I had found. Sulloway [10] showed that later-borns are more likely than first-borns to do and support radical science. I am a first-born.

Nor was self-experimentation a sufficient explanation. Self-experimentation is not new [11,12]. Almost all examples have involved dangerous drugs or medical procedures. For example, one man took arsenic to test an antidote [12]. Another man had a tooth extracted to test a new anesthetic [12]. My self-experimentation was safe. I studied treatments available to anyone. Previous self-experimentation had almost always confirmed the experimenter's beliefs. Mine often surprised me. No doubt self-experimentation was *necessary* for my progress. But its age and availability (many scientists could have done what I did) implies it was not sufficient.

My explanation of effectiveness has three parts. First (circumstances): I was in a rare and powerful position. I had the subject-matter knowledge of an insider, the freedom of an outsider, and the motivation of someone with the problem. Second (method): Self-experimentation is powerful and self-experimentation about the brain even more powerful. Third (context): Some broader ideas will make clearer why this combination of position and method worked so well.

Part 1: My powerful position

I was in an unusually good position to make progress. I had the subject-matter knowledge of an insider, the freedom of an outsider, and the motivation of someone with the problem.

The subject-matter knowledge of an insider

I was not a sleep, mood, or weight expert but I was not naive. I was a professor of psychology at a research university. From teaching introductory psychology, I had learned about sleep, mood, and weight research. I also knew about sleep research because sleep is controlled by an internal clock and my research was about a (different) internal clock [3]. My conventional research had given me a good understanding of experimental design, measurement, and data analysis.

Four examples of how my knowledge helped: 1. In 1990, unusual data analysis showed that my sleep duration had decreased a few months earlier. This finding triggered a series of events that led to the discovery that breakfast caused early awakening. 2. Because I had done research with rats, I knew about a laboratory effect called *anticipatory activity* [13]. My breakfast discovery resembled anticipatory activity in humans. This made my conclusion (breakfast causes early awakening) far more plausible. 3. From teaching introductory psychology, I knew that depression and insomnia are closely linked. The linkage made it much more plausible that something I had done to improve sleep (watch TV in the morning) improved mood. 4. My knowledge of associative learning (the main topic of animal learning, my area of expertise within psychology) made it much easier to go from Ramirez's results [9] to a new theory of weight control.

The freedom of an outsider

My self-experimentation was not my job. For a long time, I did not expect to publish it; even later, after I decided to, I did not plan to use it to gain status within a profession. This freed me to (a) do whatever worked and (b) take as long as necessary. Professional scientists cannot try anything and cannot take as long as necessary. As Dyson [14] said, "In almost all the varied walks of life, amateurs have more freedom to experiment and innovate [than professionals]."

Professional scientists are constrained in many ways. Most of their research with human and animal subjects needs to be approved by an internal review board, which may take 6 months.

Every change in protocol must be approved. Most professional scientists need grants. As Lederberg [15, p. 337] said, "Only the most accomplished and fortunate [scientist] can look beyond the renewal of their research grant." Most professional scientists need a steady stream of publications. To get tenure, you must publish a certain amount. To get your grant renewed, you must publish a certain amount. If you have graduate students, each should do publishable research. Two years to produce a paper might be okay but 10 years would be too long. It took me 10 years to begin to understand why I woke up too early. I tried one possible solution after another and eventually made progress. A professional sleep researcher could not have done this.

Professional scientists are also constrained by taboos. Alister Hardy was a biology professor at Oxford in the middle of the 1900s. Early in his career, he came up with the aquatic ape theory of human evolution. He said nothing about it, however. "I wanted to be a professor. I wanted to be a Fellow of the Royal Society," he explained [16, p. 13]. After 30 years, he had achieved these goals, so he felt free to give a talk about his theory. A journalist happened to attend and wrote about it. One of his colleagues was upset and phoned Hardy. Never *do that again*! he told him. Hardy's crazy theory had made Oxford look bad. At Hardy's memorial service, his theory was not mentioned [16, p. 14]. Because I was not a sleep, mood, or weight researcher, I did not care what they found unacceptable. Perhaps they consider self-experimentation unacceptable.

Major scientific advances are often linked to unusual freedom. Mendel's job as a monk gave him botanical freedom. His colleagues did not care about his pea plants nor what he wrote about them. Likewise, Charles Darwin could write whatever he wanted, unlike biology professors. His wealth and lack of job meant he had little to lose. Alfred Wegener, who proposed continental drift, was a meteorologist, not a geologist. His geological heresy surely did not bother his colleagues. Mendel, Darwin, and Wegener illustrate Dyson's point about the freedom of hobbyists.

The motivation of someone with the problem

I did self-experimentation to improve my own life (e.g., sleep better). In contrast, professional scientists almost never study their own problems. They have other goals.

On the face of it, professional scientists should embrace self-experimentation. It would make their life easier. No need for grants or graduate students. Not only is self-experimentation easier, it allows study of a wider range of questions. If you have tenure, you can afford to do slow, risky research. Why is long-term self-experimentation like mine not more popular?

In *The Theory of the Leisure Class* [17], Veblen argued that upper-class persons, such as professors, take considerable pains to show their social position. They do so in three ways: 1. *Display wealth*. Veblen coined the term *conspicuous consumption*. Tail fins do not improve a car's nominal function (transportation) but do show wealth. 2. *Display uselessness*. The customs of long fingernails (women) and ties (men), said Veblen, arose because both show that their possessors do not do manual labor (useful), with which long nails and ties would interfere. 3. *Display refinement*. You display refinement via activities that are conspicuous, time-consuming, and of little value. Display of useless knowledge such as "knowledge of the dead languages and the occult sciences; of correct spelling; of syntax and prosody; of the various forms of domestic music and other household art; of the latest properties of dress, furniture, and equipage; of games, sports, and fancy-bred animals, such as dogs and race-horses" [17] were some of Veblen's examples.

The last chapter of *The Theory of the Leisure Class* is about professors. As the term *ivory tower* indicates, most academic research lacks practical value (Rule 2). Several facts suggest that science professors follow Veblen's rules. Not being wealthy, they cannot

display wealth (Rule 1) but they can follow Rules 2 (uselessness) and 3 (refinement). Scientists distinguish between basic research (with no obvious value) and applied research (with obvious value); consistent with Veblen, basic research has higher status. To give a specific example, much of modern economics, especially the highly mathematical parts, has little obvious value [18]. "It's academic," a prominent economist said recently. "It is nothing like as useful to the business community as it could be" [18, p. 51]. Here *academic* means useless, as it does in the phrase *of academic value*. Although the Nobel Prize is supposed to be given for useful research, it is often given for research without clear practical value – the 2009 Medicine prize for telomere research, for example. After William Vickrey received the 1996 Nobel Prize in Economics, he told a journalist his prize-winning work was "at best … of minor significance in terms of human welfare" [18, p. 50]. According to John Cassidy, the *New Yorker* writer, the economics Nobel Prize "has fostered a professional culture that favors technical wizardry above all else" [18, p. 60] – above practical value, in particular. Veblen would say this tendency did not need much fostering. At the same time, research of great practical value, such as the discovery that smoking causes lung cancer, has not been given a Nobel Prize.

Technical wizardry without obvious value is an example of refinement. The wizardry took time to learn. In the field of statistics, professors emphasize complex numerical algorithms. Although such algorithms are much less useful than graphs [19], statistics texts are roughly 1% graphs, 99% complex numerical algorithms, reflecting how statistics professors spend their time. Likewise, scientists use unnecessary fancy words. When a scientist describes something as *rufous* rather than *brown*, he is showing refinement. Most people do not know what *rufous* means.

Veblen's ideas help explain why self-experimentation is rare among professional scientists. It violates all three of his rules. Scientists cannot display great wealth (Rule 1), but they can at least hope to get a large grant, buy expensive equipment, and have many people working for them. Because of its low cost, self-experimentation does not facilitate that. Self-experimentation such as mine is obviously useful, violating Rule 2 (uselessness). And anyone can do it, violating Rule 3 (refinement). It is *common*, in the derogatory sense.

Veblenian tendencies push scientists away from self-experimentation and away from useful work. But if you *have* a health problem, you will surely care more about alleviating it than displaying status. Some examples suggest what a difference the change in motivation makes. Paolo Zamboni's discovery of reduced blood flow in persons with multiple sclerosis [20] happened because his wife had multiple sclerosis. He was not a multiple-sclerosis expert. Dennis Mangan, a lab technician, found that mega-doses of niacin quickly cured his mother's Restless Leg Syndrome [21], something not previously reported in the scientific literature. The most important example is home blood glucose testing, an enormous advance in the management of diabetes. It was pioneered by Richard Bernstein, at the time an engineer, who had diabetes himself [22].

Part 2: Advantages of self-experimentation, especially about the brain

My self-experimentation was more powerful than conventional research for four reasons. Two came from use of self-experimentation; the other two from studying measures (sleep, mood, and weight) controlled by the brain. (Variation in weight is mainly variation in body fat, which is controlled by hunger.)

Advantages of self-experimentation

First, self-experimentation is much easier than conventional research. It allows you to test solutions to health problems much faster, more cheaply, and more flexibly. For example, self-experi-

mentation can test a new way to lose weight much faster than conventional experimentation. In [2] I reported five self-experiments about weight control. I have also done a conventional weight-control experiment with six subjects. With the time and effort it took to do that experiment, perhaps I could have done 50 self-experiments.

Second, self-experimentation measures many things at once. We continuously monitor ourselves in many ways, in the sense that our brain receives input from thousands of nerves. Our bodies can malfunction in thousands of ways. We effortlessly notice many of them. Without trying, we notice our mood, clarity of thought, sleepiness, sleep quality, coordination, hunger, thirst, most of our skin, our digestion, how our joints feel, and so on. A conventional experiment measures far less. This wide-net feature of self-experimentation was responsible for my faces/mood discovery. One morning I watched TV because I thought it might improve my sleep; the next morning, without any special preparation, I noticed my mood was much better than usual. To quantify this, let us say we monitor ourselves in 100 separate ways. A conventional experiment measures perhaps five independent dimensions. In this way self-experimentation is 20 times more powerful than conventional experimentation.

Advantages of studying the brain

Studying the brain rather than another part of the body gave me more advantages. The brain resembles a model system in two ways.

First, it changes quickly. The brain responds to external changes much faster than the rest of the body. Reaction-time experiments may put different treatments 10 s apart. In many of my rat experiments, which measured behavior, different trials involved different treatments; the trials were about a minute apart. A treatment may substantially improve brain function in hours. To substantially improve bone density might take months. Let us say the brain changes 10 times faster than other organs.

Second, it is easy to measure. The brain can be measured by measuring behavior, which I can do with my laptop. I cannot measure immune function, liver function, kidney function, bone density, heart attack risk, or dozens of other important health measures with my laptop. Perhaps the brain is 50 times easier to measure than other organs. (An exception is the skin, which is easy to measure. It makes sense that one of my examples of self-experimental power involved acne.)

The first two features make self-experimentation 1000 (50×20) more powerful than conventional research. The last two features make self-experimentation about the brain 500 (10×50) times more powerful than self-experimentation about other organs. The total improvement – self-experimentation about the brain compared to conventional research not about the brain – is 500,000 (1000×500). This means I could test 500,000 cause–effect relationships about the brain for the price that conventional researchers pay to test one cause–effect relationship about another part of the body. The combination of two big advantages (self-experimentation, brain) resembles the combination of fruit flies and salivary-gland chromosomes (much larger than other chromosomes) so helpful in the early study of genetics.

Part 3: A theory of scientific progress

Parts 1 and 2 of this explanation do not explain two features of [2]. One is the large number of accidents (completely unexpected results). There were six: 1. When I changed my breakfast from oatmeal to fruit, my early awakening got worse. 2. When I stopped eating any breakfast, my early awakening almost disappeared. 3. When I watched TV early one morning, my mood improved the next day. 4. When I stood more, my sleep improved. 5. I mysteri-

ously lost my appetite in Paris. 6. My theory of weight control, which helped me discover new ways to lose weight, was heavily based on Ramirez's accidental discovery [9]. Why so many? The other unexplained feature is the way the rate of progress increased. It took 10 years to make one practical discovery (about breakfast and sleep); in the next 12 years I made about 10 more.

To explain these features, it helps to make assumptions behind Parts 1 and 2 more explicit. A context within which Parts 1 and 2 make sense has four assumptions.

Assumption 1: Scientific progress has a power-law-like distribution

Progress in science, as far as I can tell, is a mix of many tiny steps and a few big steps. When I have looked closely at big advances in understanding that led to practical discoveries, they were always built on a great deal of data. Each bit of data helped but there was great variation in how much. Almost all the observations were routine. They were unsurprising; they confirmed an idea already quite plausible. They repeated a well-known effect, for example. A small number were not routine; in some way they contradicted what was expected. Of that small number, most showed that a plausible idea was wrong. The Michaelson–Morley experiment, which disproved the ether theory, is an example. A small fraction of the small number immediately suggested a new idea. For example, the Geiger–Marsden experiment found that a few subatomic particles aimed at gold foil bounced backwards. It suggested a new theory of the structure of matter.

My own research resembled the historical examples. My discovery that breakfast caused early awakening (Example 1 in [2]) came from 10 years of trial and error. During those 10 years, I recorded my sleep most days. The sleep data I collected can be divided into three groups: (a) *Tiny progress*. I woke up too early thousands of times. In almost all cases, this was unsurprising but nevertheless contributed a tiny bit of progress, like a control group. (b) *Medium-sized progress*. Sometimes I tried a new solution to the problem – more exercise, for example. All my attempted solutions failed. That data contributed more than a tiny bit of progress; they showed that plausible beliefs were wrong. (c) *Large progress*. Two data sets were big steps forward: 1. When I started eating fruit for breakfast, early awakening increased. 2. When I stopped eating breakfast, early awakening nearly vanished.

These observations suggest that the distribution of scientific progress resembles a power-law (Pareto) distribution, which is linear on log–log coordinates (Fig. 1). You sample from the distribution when you gather data (everyday life, surveys, experiments). Most increases in knowledge are tiny, a tiny fraction are large. The largest ones are often called accidental, but Fig. 1 implies this is misleading. The simplicity of the function implies that all progress large (accidental), medium, or small comes from the same underlying process.
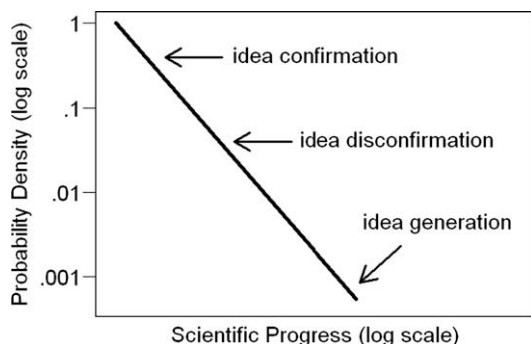
Specific evidence for such a distribution is, in addition to [2], the many medical examples described by Meyer [23]. Each of Meyer's examples of medical progress involved many small steps and one large step. More evidence is the large number of non-medical examples where scientific progress came from an accident. The *Wikipedia* entry for *serendipity* lists 44 examples from biology, chemistry, physics, and astronomy. (Plus 15 from pharmacology, three from medicine not involving drugs, and nine from engineering.) For example, the first battery was made soon after Galvani noticed that a spark from a metal scalpel caused a dead frog's leg to twitch – an accidental discovery. (And another example of the advantages of studying neurons.) Studying this phenomenon, he found that a frog's muscles would twitch when it was in contact with two different metals. The first battery contained layers of different metals. A different sort of support for Fig. 1 is that the number of times a scientific paper is cited has a power-law distribution [24].

"Genius is one percent inspiration and ninety-nine percent perspiration," said Thomas Edison. Fig. 1 says you need perspiration (a lot of sampling) to find inspiration (a big step forward). Tukey [25] distinguished exploratory data analysis (e.g., graphs) and confirmatory data analysis (e.g., *t* tests). Again, Fig. 1 says such distinctions are misleading. It says you generate ideas the same way you test them. For years, like Tukey, I believed that idea testing and idea generation required different methods. The title of [2] ("Self-experimentation as a source of new ideas") reflects this distinction. It says self-experimentation is a good way to generate ideas – as if it were *not* a good way to test them. Table 1 of [2] says the same thing. Fig. 1 says I was wrong. The facts support Fig. 1. I used self-experimentation to test ideas many times. Most of the idea-generating accidents came when I was testing an idea. Ramirez's surprising observations [9] happened while testing an idea.

Assumption 2: The slope varies

The second assumption is that the slope of the progress distribution varies. It depends on what you know and do. Subject-matter knowledge flattens the distribution (left panel of Fig. 2). A sleep expert will learn more from one sleep datum than a non-expert. Pasteur's "chance favors the prepared mind" is a subset of the left panel of Fig. 2, which says all data favor the prepared mind. In addition, scientific methods flatten the distribution. An experiment will produce more progress than the same amount of non-experimental data (right panel of Fig. 2). Likewise, better data analysis flattens the distribution. Such shifts of power-law-like distributions have been observed [4].

The idea that knowledge improves the slope explains why my rate of progress increased after 10 years. After 10 years of trial and error, I discovered that breakfast caused early awakening. This was a big step forward. I believe it improved the slope of my pro-
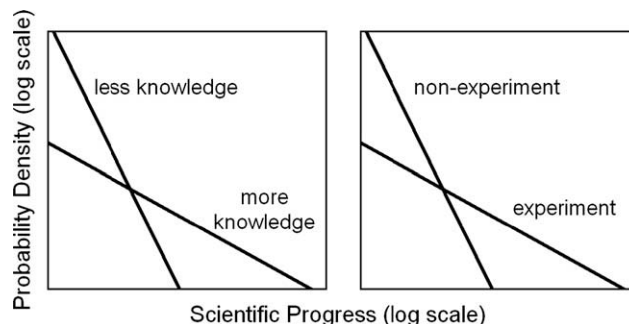

Fig. 1. The distribution of scientific progress. Both axes are log-transformed. Almost all advances are very small; a tiny fraction are very large. Data that confirms an already-plausible idea is a small advance; data that disconfirms a plausible idea is a medium-sized advance; data that generates a new idea is a large advance.

Fig. 2. Subject-matter knowledge (left panel) and research type control the slope of the distribution of progress.

gress distribution because it increased my knowledge. It suggested that other powerful and beneficial treatments might be found among the elements of Stone-Age life. This idea led to three more discoveries (morning faces and mood, standing and sleep, sleep and colds).

Assumption 3: Sampling rate varies

The third assumption is that the rate of sampling from the distribution varies. Some methods allow more sampling than others. For example, a fast experimental design (e.g., a simple design) allows a higher sampling rate than a slow design (e.g., a more careful one). This might seem obvious, but in what I have read about experimental design sampling rate was never mentioned.

Assumptions 1–3 explain much of why my self-experimentation was unusually effective. The experts – sleep experts, for example – know more about the subject, use better equipment, and have more subjects per experiment than I do, so their distribution is flatter than mine. But their slope advantage is overwhelmed by my advantage in sampling rate.

Assumption 4: Freedom and motivation matter

Earlier I gave six examples (Mendel, Darwin, Wegener, Zamboni, Mangan, Bernstein) that suggested the importance of freedom and motivation. They were inside/outside correlations: Outsiders with unusual motivation or freedom made more progress than insiders. Within-profession correlations also exist, where insiders with unusual freedom or motivation make more progress than other insiders. In *The New Yorker* in 2006–7, Atul Gawande wrote two articles about medical innovation [26,27]. The first [26] was about Apgar scores, which describe the health of a newborn baby. Since their introduction in 1953, Apgar scores have gradually improved, saving thousands of lives, even though, according to Gawande, obstetricians do not do research in the approved "evidence-based medicine" way. In addition, obstetricians are lower-status: "Doctors in other fields have always looked down their masked noses on their obstetrical colleagues," wrote Gawande. Both features – less "correct" research methods and lower-status – suggest more freedom than usual (more freedom of research method) and unusual motivation (less desire for status because obstetricians chose a lower-status job within medicine). Gawande's second article [27] was about the use of checklists to reduce surgical errors, which was pioneered by a doctor named Peter Pronovost. Asked why he did this work, Pronovost said his father had died from a medical error [28], an unusual motivation. Gawande did not mention this but did note the strangeness of Provonost's work: no "multimillion-dollar grant", no "swarm of doctoral students and lab animals" [26] – that is, no conspicuous wealth. "He's focussed on work that is not normally considered a significant contribution in academic medicine," wrote Gawande. "Yet his work has already saved more lives than that of any laboratory scientist in the past decade" [26]. Free from Veblenian tendencies, Provonost achieved highly-useful results.

A test of this explanation

In the previous sections, I have described the explanation and simple evidence for it. The rest of this article describes more complicated support (this section), related work (next section), and a broader but compatible explanation ("Science and Human Nature").

A philosophical idea called Reichenbach's Common Cause Principle helps explain rare events. It says if two events are correlated, either one causes the other or they have a common cause. *Have a common cause* means if you look at the sequences of trigger events (A caused B caused C) that led to each of the two events, you should find a common element from which the two sequences branch. For example, suppose Events A and B are correlated. Then the sequence of events that led to A (X caused Y caused A), if extended back in time, should meet the sequence of events that led to B (X caused Z caused B). In this example, the common cause (trigger event) of A and B is X. In practice, Reichenbach's Principle amounts to *lightning doesn't strike twice in one place for different reasons* [29]. For example, one night you hear a strange sound downstairs (Rare Event 1). In the morning, your TV is gone (Rare Event 2). Rare Events 1 and 2, correlated in space and time, surely had a common cause (a burglar). Rare Event 2 helps choose between possible explanations of Rare Event 1. The research I want to explain [2] had several rare features:

1. *Great novelty*. No one had previously concluded anything close to the ideas that breakfast causes early awakening, standing reduces early awakening, morning faces improve mood a day later, or sugar water causes weight loss.
2. *Diversity*. The conclusions involved areas of research – sleep, mood, and weight – usually studied by different researchers. Sleep researchers do not study weight, for example.
3. *Long-term self-experimentation*. No other published self-experimentation has lasted long as mine (12 years). Most published self-experimentation has lasted a few weeks or less [12].
4. *Low cost*. It cost almost nothing.
5. *Not an expert*. None of my previous publications were about sleep, mood, or weight.
6. *Personal*. I studied my own problems.
7. *Everyday treatments*. The treatments used in my experiments were variations of everyday life, such as not eating breakfast, standing more, seeing faces, and eating ordinary food.
8. *Low publication rate*. Only two scientific articles [2,7] have come from this research. Two papers in 13 years is a low rate of publication.
9. *Immediate use*. The conclusions were helpful right away. I slept better, was in a better mood, and lost weight. The Shangri-La Diet [30], based on the weight-control conclusions, was published only 2 years after [2] and has helped many people lose weight (see boards.shangriladiet.com for examples).

Reichenbach's Principle says these nine rare features should have a common cause.

The explanation I have given here passes this test. The triggering event for all nine features was that I started self-experimentation involving the brain. I had near-expert subject-matter knowledge yet my motivation was unusual: personal benefit. So I (a) had a favorable distribution of progress function, (b) could sample from it easily and often, and (c) could try anything. This combination was so potent that it led me to ideas of great novelty (Feature 1) in several areas (Feature 2). Because it worked so well, I did it for a long time (Feature 3), nothing expensive was needed (Feature 4), and I did not need to be an expert (Feature 5). Because I was doing it for personal benefit (Feature 6), I studied everyday treatments (Feature 7). Because I was not doing it for career reasons, I did not publish much (Feature 8). Because I was studying myself, the findings had immediate use (Feature 9).

Other explanations of the effectiveness fail this test in the sense that they do not explain all nine features. The explanation that I was smart does not explain the personal benefit (Feature 6), everyday treatments (Feature 7), low publication rate (Feature 8), and immediate use (Feature 9). The explanation that I did self-experimentation fails to explain the everyday treatments (Feature 7) and low publication rate (Feature 8). The explanation that I was an outsider fails to explain the personal benefits (Feature 6) and immediate use (Feature 9).

Related work

The idea that outsiders have more freedom than insiders has appeared in several forms. Veblen argued that the freedom enjoyed by outsiders gives them a problem-solving advantage [31]. Sulloway argued that later-borns are more likely to support radical science than first-borns because they are less invested in the status quo [10]. An experiment in an American lab found that students did better on a problem-solving task when told the problem came from Greece than when told it came from nearby [32]. The students knew less about Greece than America, including less about constraints in Greece. Maybe this helped them think of more solutions. Likewise, outsiders know less about constraints than insiders.

The idea that people devote considerable resources to displaying status is common within anthropology. A recent example is *Watching the English: The Hidden Rules of English Behaviour* [33], which includes a lot about class display. After Veblen, the term *status symbol* became common.

Power-law-like distributions have been observed in many situations [34,35]. The theorists Zipf [36], Mandelbrot [37], and Bak [38] used them as a unifying theme. My use of power-law-like distributions is closest to Taleb's [39], who emphasized their implications for everyday life. Taleb argued that we are poor at anticipating extreme events – we "under-expect" them. Taleb mainly discussed financial bad news, such as stock market crashes, but his ideas also apply to good news, including scientific discovery. He pointed out to me the connection between his work and mine. Just as we under-expect extreme bad news, we under-expect extreme *good* news. Financiers take too many risks; scientists do not take enough. Self-experimentation makes it easy to test long-shot ideas.

Science and human nature

My explanation of the effectiveness of my self-experimentation can be summed up by saying I was in an unusual position and picked up an unusual tool. One rare event (unusual position plus unusual tool) caused another (unusual results). My work looks unusual, yes. If you read 1000 scientific articles, including mine, mine would stand out. But from another point of view, the *other* 999 articles are the outliers. Like most human activity, my self-experimentation fit well with human nature. Professional science – represented by those 999 articles – does not. This is another reason, I believe, that my self-experimentation was unusually effective: It resembled time-tested ways of exploring.

Science resembles part of human nature: the way we learn by doing. My first self-experiments were inspired by an article about how to teach math [40]. "The best way to learn is to do," it said [40, p. 466]. The author meant this as a statement about human nature. Learning by doing, which learning researchers call *instrumental learning*, is easy to study in animals such as rats and birds. This implies that our brains have been learning by doing for a long time. A scientific experiment is learning by doing assisted by wisdom about method (e.g., experimental design). In my self-experimentation, I used scientific methods to help me learn by doing about personal problems, such as how to sleep better. The methods helped me do something (learn by doing) I was predisposed to do.

Some jobs fit human nature well. They take advantage of natural tendencies and provide plenty of what we need to be happy. Other jobs do not. They make workers act against natural tendencies or do not fulfill basic needs. As a full-time job, science is a poor fit. One reason I discussed earlier: Desire to display status interferes with progress. Another reason derives from the place of exploration in our evolutionary past. A few million years ago, when the human lineage split from other primate lineages, our ancestors were foragers. Foraging includes exploitation and exploration. Sometimes you find food by returning to places where you have already found it (exploitation); sometimes you look in new places (exploration). Ants show the exploitation/exploration difference clearly. Sometimes they follow other ants to a food source (exploitation), making a line on the kitchen floor; sometimes they wander alone (exploration).

Humans are occupational specialists, not foragers. If you randomly select 100 people, they are likely to have 100 different jobs. No other species is like this. I have proposed that our brains changed in many ways to make occupational specialization possible [41]. During the million-year shift from foraging to occupation, however, I believe some things stayed the same. The constant elements caused my self-experimentation to be more effective than professional science.

Pure foraging, I proposed, was followed by an age of hobbies, during which small amounts of time were devoted to hobby-like activities, presumably tool-making. At first, these hobbies were all exploration – there was no accumulated knowledge to exploit. Specialization began: Different people had different hobbies, just as now. By trial and error (hobbyist exploration), knowledge about where to find and how to control materials slowly accumulated. It was passed on by imitation. Hobbyist exploration, as far as I can tell, had four properties:

- *By specialists*. How to make better baskets was figured out by those whose hobby it was. "The best way to learn is to do" and they were the ones doing.
- *Surrounded by exploitation*. Hobbies could be pure exploration but, like today's hobbies, were done in your spare time. During the rest of the day, a hobbyist foraged, which was mostly exploitation.
- *Benefits of discoveries went to the explorer*. If a knife maker discovered how to make better knives, he would benefit. Veblen called the desire to make things well "the instinct of workmanship" [42]. Perhaps it comes from the pleasure we get from well-made things.
- *Immediate benefit*. The discoveries were beneficial right away.

The last three properties resembled the exploration that is part of foraging. Foraging does not involve specialization (all members of a species search for and eat the same food) but, like hobbyist exploration, foraging exploration is surrounded by exploitation (animals do not take long breaks to explore), the benefits flow to the discoverer (you eat what you find) and the benefits are immediate (you usually eat it immediately).

As hobbies became more skilled and their products more useful, trading began. A knife specialist traded a knife he had made for a spear made by a spear specialist. Both benefited from the trade. Think of the tool-making knowledge available to each person as a pile. Trade made the pile more valuable. When the pile and its value grew large enough, part-time jobs became possible: You could trade for some of your needs. Eventually full-time jobs became possible. Each person learned a small piece of the pile to make a living.

After the transition from hobbies to jobs, the pile of vocational knowledge continued to grow by what I will call *artisanal* exploration. People with various specialties slowly improved the state of their art. Artisanal exploration had the same properties as hobbyist exploration. Again, it was by specialists. Those who made baskets for a living figured out how to make better baskets. For a non-expert to try to improve basket-making would have been a waste of time. The experts had a huge head start. Again, the exploration was surrounded by exploitation – now, exploitation of specialized knowledge to make a living. Basket-makers did not spend long

Table 1
Exploration in different contexts.

Property	Type of exploration				
	Foraging	Hobbyist	Artisanal	Professional science	My self-experimentation
By experts	No	Yes	Yes	Yes	Yes
Mixed with exploration	Yes	Yes	Yes	No	Yes
Explorers benefit	Yes	Yes	Yes	No	Yes
Immediate benefit	Yes	Yes	Yes	No	Yes

periods of time doing pure research, judging by modern-day artisans, e.g., [43]. They tried new ways of doing things in the midst of their usual work. Again, the benefits from successful exploration went to the discoverer. A basket-maker who discovered how to make better baskets became able to make baskets that would fetch a higher price. Again, the benefits were immediate. If a basket-maker found a better way to make baskets, he could immediately make better baskets.

These properties make sense. Property 1 (by experts): Experts take advantage of the many experts that have preceded them (their teachers). They can focus their exploration to areas more likely to pay off. A non-expert would not know where to begin. Property 2 (surrounded by exploitation): A great deal of artisanal research may have been triggered by accidental observations: Something came out better or worse than usual. This led to research about why. The starting point, the accidental discovery, cost nothing because it happened during normal work. The subsequent research into cause was likely to pay for itself by revealing something new that made a difference. A modern example is that a doctor who regularly sees patients (exploitation) is more likely to know what research will be most beneficial than a doctor who never sees patients – who only does research. Properties 3 (immediate benefits) and 4 (benefits to the discoverer) motivated and sustained the exploration. To a hobbyist or artisan, Properties 3 and 4 would seem essential. Why explore if you do not reap the benefits?

My self-experimentation had all four properties; professional science has only the first. *Property 1* (by experts): I was not exactly a sleep expert (or mood expert or weight expert) but I was close to being one. Professional science is done by experts. *Property 2* (surrounded by exploitation): My self-experimentation was a small part of what I did. Most of my time was spent exploiting my expertise in experimental psychology by doing mainstream research and teaching. For a professional scientist, who spends all his time doing science, there is no separate exploitation. A sleep researcher does not make a living giving sleep advice, for example. Most university professors teach, but at research universities teaching is a small part of the job. *Properties 3* (immediate benefit) *and 4* (benefit to explorer): When I discovered how to sleep better, it enabled me to sleep better right away. Likewise with my other discoveries. Professional scientists, as I have said, do not study their own problems. Almost none of the thousands of scientific papers published every year has immediate benefit. Of the tiny fraction with immediate benefit, almost none benefits the scientists involved (in the sense of improving their everyday life).

Table 1 summarizes these comparisons. My self-experimentation resembled foraging, hobbyist, and artisanal exploration, Professional science is a poor match for any of them. The similarity of foraging, hobbyist, and artisanal exploration suggests that our brains are well-suited for jobs with a lot of exploitation and a little exploration. Although full-time scientists are expected to explore full-time, full-time exploration is very uncomfortable. Imagine being hungry and having no idea where to find food. Imagine going door-to-door, asking strangers for donations, all day, every day. Or calling numbers out of phone book, soliciting business. Science requires freedom. But, with the freedom given them, professional scientists turn their job into one that is primarily exploitation. They

choose to do research that will generate a steady stream of scientific articles. They do so because they want steady visible progress, which exploitation can supply but exploration cannot. In addition, they use their freedom to display status. Science demands more freedom than other jobs. Almost all current jobs (bus driver, factory worker, secretary, policeman, graphic designer) are almost all, or entirely, exploitation. The job-holder uses specialized skills to do the job. The amount of time spent exploring is small or zero. The job of professional scientist is the only exception, the only job I know of where exploration is supposedly the main activity. In contrast, my self-experimentation was exploration consistent with human nature. I did not require status and steady progress because I got them from my job (professor). I did not need to shape my exploration so that it provided them.

The difference between exploitation and exploration is the difference between low-risk/low-payoff activities (exploitation) and high-risk/high-payoff activities (exploration). Foragers and artisans invest almost all their time in low-risk activities. I could invest a small fraction of my time in high-risk self-experimentation because the rest of my job was low-risk. Professional scientists, reasonably enough, want to invest almost all their time in low-risk activities, such as experiments that are likely to work. So a large fraction of their research – say, 95% – plays it safe and is unlikely to yield the sort of discoveries I made. What about the remaining 5%? Surely scientists do *some* high-risk/high-reward research. I think the problem lies with Properties 3 and 4. There is no point doing high-risk/high-reward research unless the scientist himself will benefit. To a scientist, high-risk/high-reward research is not research with practical value (such as how to sleep better). It is research that will win a Nobel Prize. Useful research is low-status. Because the Nobel Prize must be high-status (the winner is selected by high-status scientists), useful research tends to be excluded. So 95% of a professional scientist's research does little good for one reason (desire for steady progress), the remaining 5% for another reason (high-status = useless).

When science and human nature fit together (my self-experimentation), progress is fast; when they do not (conventional research), progress is slow. One reason my self-experimentation was surprisingly effective is that, compared to putting square pegs in round holes, putting round pegs in round holes is surprisingly easy.

Conflict of interest statement

None.

References

[1] Wigner E. The unreasonable effectiveness of mathematics in the natural sciences, Commun Pure Appl Math 1960;13(1). Retrieved from: <http://www.dartmouth.edu/~matc/MathDrama/reading/Wigner.html> [13.04.2010].

[2] Roberts S. Self-experimentation as a source of new ideas: ten examples about sleep, health, weight, and mood. Behav Brain Sci 2004;27:227–88. Retrieved from: <http://escholarship.org/uc/item/2xc2h866> [13.04.2010].

[3] Roberts S. Isolation of an internal clock. J Exp Psychol Anim Behav Process 1981;7:242–68.

[4] Gharib A, Derby S, Roberts S. Timing and the control of variation. J Exp Psychol Anim Behav Process 1981;27:165–78.

[5] Roberts S. Surprises from self-experimentation: sleep, mood, and weight. Chance 2001;14(2):7–18. Retrieved from: <http://escholarship.org/uc/item/5bv8c7p3> [05.04.2010].

[6] Eady EA, Cove JH, Blake J, Holland KT, Cunliffe WJ. Recalcitrant acne vulgaris. Clinical, biochemical and microbiological investigation of patients not responding to antibiotic treatment. Br J Dermatol 1988;118:415–23.

[7] Espersen F. Resistance to antibiotics used in dermatological practice. Br J Dermatol 1998;139(Suppl. 53):4–8.

[8] Kuhn TS. The structure of scientific revolutions. Chicago: University of Chicago Press; 1962.

[9] Ramirez I. Stimulation of energy intake and growth by saccharine in rats. J Nutr 1990;120:123–33.

[10] Sulloway F. Born to rebel: birth order, family dynamics, and revolutionary genius. New York: Pantheon; 1996.

[11] Altman LK. Who goes first? The story of self-experimentation in medicine. New York: Random House; 1987.

[12] Fiks AR. Self-experimenters: sources for study. Westport, CT: Praeger; 2003.

[13] Mistlberger RE. Food-anticipatory circadian rhythms: concepts and methods. Eur J Neurosci 2009;30:1718–29.

[14] Dyson F. In praise of amateurs. New York Rev Books 2002;49(19). Retrieved from: <http://www.nybooks.com/articles/15870> [23.03.2010].

[15] Lederberg J. Does scientific progress come from projects or people? Curr. Contents 1989;12(48):336–44.

[16] Morgan E. The naked Darwinist. Eildon Press; 2008.

[17] Veblen T. The theory of the leisure class. New York: Macmillan; 1899. Retrieved from: http://www.geocities.ws/veblenite/txt/tlc.txt [13.04.2010].

[18] Cassidy J. The decline of economics. The New Yorker 1996(December 2):50–60.

[19] Roberts S. Plot your data. Nutrition 2009;25:608–11. Retrieved from: <http://sethroberts.net/articles/2009%20Plot%20your%20data.pdf> [13.04.2010].

[20] Singh AV, Zamboni P. Anomalous venous blood flow and iron deposition in multiple sclerosis. J Cerebr Blood Flow Metab 2009;29:867–78.

[21] Mangan D. A case report of niacin in the treatment of restless legs syndrome. Med Hypotheses 2009;73:1072.

[22] Bernstein R. Dr. Bernstein's diabetes solution: the complete guide to achieving normal blood sugars revised & updated. New York: Little, Brown; 2003.

[23] Meyers MA. Happy accidents: serendipity in modern medical research. New York: Arcade Publishing; 2007.

[24] Gupta HM, Campanha JR, Pesce RAG. Power-law distribution for the citation index of scientific publications and scientists. Braz J Phys 2005;35:981–86. Retrieved from: <http://www.sbfisica.org.br/bjp/files/v35_981.pdf> [13.04.2010].

[25] Tukey JW. We need both exploratory and confirmatory. Am Statist 1980;34:23–25. Retrieved from: <http://www.ece.rice.edu/~fk1/classes/ELEC697/TukeyEDA.pdf> [13.04.2010].

[26] Gawande A. The score. The New Yorker 2006;(October 9):58–67. Retrieved from: <http://www.newyorker.com/archive/2006/10/09/061009fa_fact> [13.04.2010].

[27] Gawande A. The checklist. The New Yorker 2007;(December 10);86–95. Retrieved from: <http://www.newyorker.com/reporting/2007/12/10/071210fa_fact_gawande> [13.04.2010].

[28] Dreifus C. Doctor leads quest for safer ways to care for patients. New York Times 2010;(March 9):D2. Retrieved from: <http://www.nytimes.com/2010/03/09/science/09conv.html> [13.04.2010].

[29] Roberts S. Evidence for distinct serial processes in animals: the multiplicative-factors method. Anim Learn Behav 1987;15:135–73.

[30] Roberts S. The Shangri-La diet. New York: Putnam; 2006.

[31] Veblen T. The intellectual pre-eminence of Jews in modern Europe. Political Sci Quart 1919;34(1):33–42.

[32] Jia L, Hirt ER, Karpen SC. Lessons from a faraway land: the effect of spatial distance on creative cognition. J Expt Soc Psychol 2009;49:1127–35.

[33] Fox K. Watching the English: the hidden rules of English behaviour. London: Hodder & Stoughton; 2005.

[34] Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. SIAM Rev 2009;51:661–703. Retrieved from: <http://arxiv.org/PS_cache/arxiv/pdf/0706/0706.1062v2.pdf> [13.04.2010].

[35] Buchanan M. Ubiquity: the science of history ... or why the world is simpler than we think. New York: Crown; 2001.

[36] Zipf GJ. Human behavior and the principle of least effort: an introduction to human ecology. Cambridge MA: Addison-Wesley; 1949.

[37] Mandelbrot B. New methods of statistical economics, revisited: short vs. long tails, Gaussian vs. power-law distribution. Complexity 2009;14:5–65. Retrieved from: <http://www.math.yale.edu/mandelbrot/web_pdfs/ComplexityNew.pdf> [13.04.2010].

[38] Bak P. How nature works: the science of self-organized criticality. New York: Springer-Verlag; 1996.

[39] Taleb N. The black swan: the impact of the highly-improbable. New York: Random House; 2007.

[40] Halmos P. The problem of learning to teach: the teaching of problem solving. Am Math Month 1975;82:466–70.

[41] Roberts S. Diversity in learning. Ideas That Matter 2005;3(3)39–43. Retrieved from: <http://www.sethroberts.net/about/2005_diversityinlearning.pdf> [15.04.2010].

[42] Veblen T. The instinct of workmanship. New York: Macmillan; 1914. Retrieved from: <http://www.archive.org/details/instinctofworkma00vebl> [15.04.2010].

[43] Bilger B. A better brew. The New Yorker 2008;(November 24):86–99. Retrieved from: <http://www.newyorker.com/reporting/2008/11/24/081124fa_fact_bilger> [15.04.2010].