



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Journal of Business Research



## The reception of my self-experimentation

Seth Roberts\*

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

## ARTICLE INFO

## Article history:

Received 1 June 2010  
Received in revised form 1 November 2010  
Accepted 1 February 2011  
Available online 12 March 2011

## Keywords:

Self-measurement  
Introspection  
Self-tracking  
Sleep  
Mood  
Innovation

## ABSTRACT

Self-experimentation makes some experiments much easier. They might be impossible without it. It can generate plausible new ideas by (a) producing surprising results, which suggest new ideas; and (b) allowing implausible ideas to be cheaply tested. For example, one of my self-experiments showed that seeing faces in the morning raised my mood the next day. Another found that standing more than 8 h while awake made me sleep better. A long article about my self-experimentation (Roberts, 2004) got a chilly reception within my department (psychology). It got a much better reception elsewhere. Blog posts about it led to a popular book (Roberts, 2006) based on one of the results. My self-experimentation combined insider knowledge, outsider freedom, and the motivation of someone who personally benefits from the research – a potent combination.

© 2011 Elsevier Inc. All rights reserved.

## 1. Introduction

Researcher introspection and my self-experimentation (Roberts, 2004) share something important: their reason for existence. Researcher introspection makes it easier to study hard-to-study topics. Gould (1991) used it to study how various products changed his “vital energy” (Gould, 1991, p. 194), meaning how focused he felt, how energetic he felt, and so on. “Much of consumer research [using conventional methods] has failed to describe many aspects of my own consumer behavior,” he wrote (Gould, 1991, p. 194). He used introspection to fill the gaps and described in detail experiences over many years. To collect such data in other ways would have been impossible. Earl (2001) used his own experience to shed light on why people pay high prices to attend rock concerts when recordings are cheap. The question would have been much harder to answer in other ways. Hirschman (1990, p. 115) described how a “near-death experience altered her perception of consumption [and] consumer behavior research.” To learn from near-death experiences using common research methods is probably impossible. Reid and Brown (1996) used introspection to argue that shoppers classified *apathetic* experience a lot of emotion during shopping. It would have been much harder to gather similar data from others.

To say researcher introspection should not be used because of flaws (Wallendorf and Brucks, 1993) ignores this. “One might as well...jerk the Cuisinart because it won't chop firewood” (McKibben, 1983, p. 44). All

methods have flaws and limitations. Reasonable criticism of an unusual method is to argue in specific cases – specific uses of the method – that the conclusions were false or misleading. The main critique of researcher introspection (Wallendorf and Brucks, 1993) did not do this.

My self-experimentation (Roberts, 2004) filled a similar niche: It made difficult or impossible experiments much easier. The otherwise-impossible experiments (a) produced surprising results that generated new ideas and (b) tested ideas I couldn't have tested otherwise.

## 2. Background

My self-experimentation began when I was a graduate student in psychology. I wanted to learn how to do experiments. I believed “the best way to learn is to do” (Halmos, 1975, p. 466) so I tried to do as many experiments as possible. Self-experiments were much faster than the rat experiments I'd been doing.

One of my first self-experiments tested two acne medicines (tetracycline and benzoyl peroxide) my dermatologist had prescribed. In the beginning, I thought tetracycline worked and benzoyl peroxide did not. In a few months, my results showed the opposite: benzoyl peroxide worked and tetracycline did not. I was stunned. It had been so easy to learn something surprising and useful.

A few years later, I started waking up too early. Because of my acne experience, I began to self-experiment (record my sleep and try various treatments) to improve the situation. I tested a wide range of non-drug treatments (e.g., exercise, eating cheese). For ten years, nothing worked. Nothing I tried made a difference. Then something curious happened (Example 1 of Roberts, 2004). My sleep records showed that at the same time I had lost weight I had started to sleep

\* 1152 Arch Street, Berkeley, CA 94708, United States.  
E-mail address: twoutopias@gmail.com.

less. I showed this to my students. A week later, one of them came to my office and suggested a diet with high water content. It would make me lose weight and need less sleep, he said. I tried it. Nothing happened. The next time I saw the student, I told him that the water diet had made no difference. “How much fruit are you eating?” he asked. “Four pieces a day,” I said. He said he ate six pieces a day. So I changed my breakfast from oatmeal to two pieces of fruit. As soon as I made the change, my early awakening got worse. Apparently breakfast mattered. I tried several different breakfasts. Eventually I concluded that any breakfast with calories caused early awakening. This made more sense than you might think. If a mammal is fed at the same time every day, it will become active a few hours earlier, a phenomenon called *anticipatory activity*. I'd been eating breakfast at 7 am and waking up at 4 am.

These results also suggested a broader point. Our bodies were shaped by evolution to work well under Stone-Age conditions. Stone-Age people didn't eat breakfast. So it made some sense that breakfast caused trouble. My breakfast results made me think perhaps many health problems were due to differences between modern life and Stone-Age life.

After that, I focused on Stone-Age/modern differences. *Can this or that element of Stone-Age life improve health?* I asked. This seemed to be a good approach because my rate of discovery increased. Soon after the breakfast results, I used an idea about Stone-Age life to try to improve my sleep (Example 2 of Roberts, 2004). Even after I stopped eating breakfast, I continued to wake up too early. Apparently breakfast was not my only problem. What else about my life might be causing trouble? I knew that conventional experiments had shown that social contact controls when we are awake: We tend to be awake at the times we have social contact. I believed that Stone-Age people had plenty of social contact in the morning, whereas I lived alone. Putting these two things – the research result and my idea about Stone Age life – together suggested this: Perhaps lack of human contact in the morning made my sleep worse. Data also suggested that TV could have the same effect as social contact. Perhaps if I watched TV early in the morning, it would improve my sleep.

One morning I tried it. Nothing happened – or so it seemed. After watching the TV, I felt no different than usual. However, when I awoke the next morning, I felt remarkably good – cheerful, calm, yet energetic. I couldn't remember ever feeling so good that early in the morning. I studied the situation further and, to my astonishment, figured out that if I saw faces on TV early in the morning it raised my mood the *next* day – not the same day. If I saw faces Monday morning, I felt better on *Tuesday*.

I did a simple experiment that showed the effect and helped explain it. On some days I saw about 60 min of faces on TV starting at 6 am; on other days everything was the same except the faces were covered. In addition, I rated my mood every few hours. I used three mood scales. On Scale 1, I rated myself on the dimension sad/happy; on Scale 2, on the dimension reluctant/eager; and on Scale 3, on the dimension irritable/serene. Each scale went from 0 to 100, with 50 = neutral, 60 = slightly positive, 70 = somewhat positive, and 80 = quite positive. For example, if I felt slightly happy I would rate myself 60 on the sad/happy dimension. If I felt slightly sad, 40 on that dimension. If I felt somewhat eager, I would rate myself 70 on the reluctant/eager dimension. To get an overall score, I averaged the three scales.

Fig. 1 shows the results. Each point is a different measurement. The upper panel, which shows my mood at 4 pm, shows the next-day effect I'd noticed. The 6-am faces greatly improved on my mood, but with a one-day lag. Several things made this unlikely to be a placebo effect. First, the initial observation was a huge surprise. Second, the size of the effect depended on many details. For example, East/West travel across many time zones eliminated the effect for weeks. Third, experiments about the effect often produced results often different from what I had expected. Fourth, other results, where a placebo effect was impossible, pointed in the same direction.

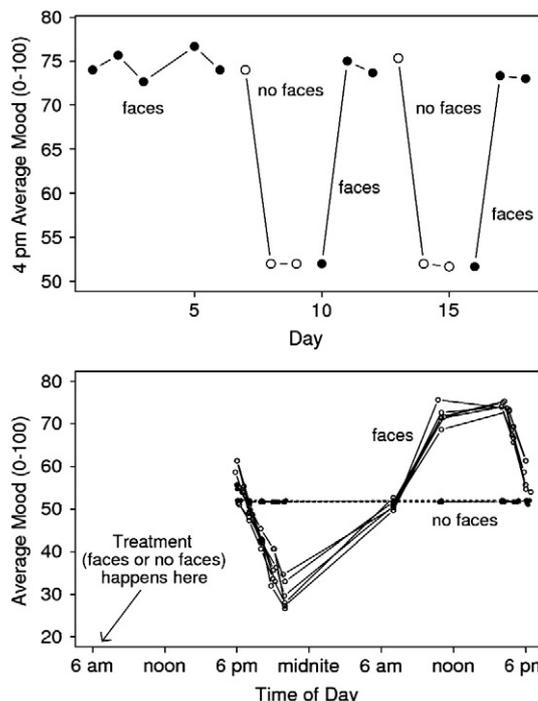


Fig. 1. Mood ratings over 17 days. Upper panel: mood at 4 pm. Lower panel: time course of the effect. In both panels, each point is an average over three ratings, one from an unhappy/happy scale, one from an irritable/serene scale, and one from a reluctant/eager scale. In the lower panel, each line is a separate day. The data start about 12 h after the treatment because that's when the treatment began to make a difference.

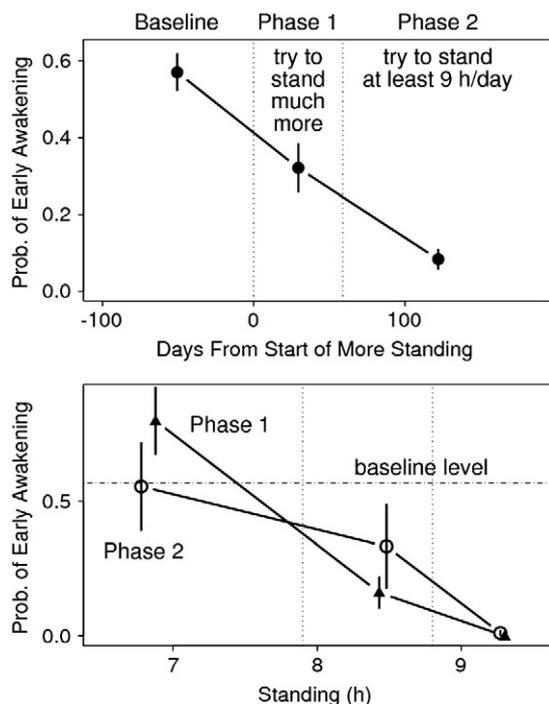
The lower panel of Fig. 1 shows how my mood changed throughout the day. It shows that the 6-am faces caused an oscillation in mood (low at night, high during the day) that started about 6 pm of the day I saw the faces – in other words, about 12 h after I saw them – and lasted about 24 h. These results suggest (a) we have an oscillator that controls our mood and (b) that oscillator is given a “push” (caused to oscillate) by seeing faces in the morning. See Roberts (2004) for more about this.

Another Stone-Age-influenced discovery was that standing improved my sleep (Example 3 of Roberts, 2004). One day a colleague said it would be nice if typing counted as exercise. Her remark made me wonder if standing counted as exercise – that is, resembled conventional exercise. If I stood a lot, would I lose weight? My belief that Stone-Age life contained elements crucial for health made this question interesting to me. Surely Stone-Age people stood much more than I did.

I worked, ate, and talked on the phone standing up. My weight didn't change. However, to my surprise, my sleep records showed that I was waking up early less often. (Aerobic exercise hadn't helped.) Fig. 2 shows what happened. I defined an instance of early awakening to be a morning when I fell back asleep between 15 min and 6 h after waking up for the first time. (For example, wake up at 4 am, fall back asleep at 7 am.) The top panel of Fig. 2 shows how the probability of early awakening changed over time. During a baseline period (before lots of standing), I woke up early about 60% of mornings. During the first period I tried to stand more (Phase 1), I woke up early about 30% of mornings.

When I analyzed the data from Phase 1, I saw that standing seemed to have no effect unless I stood 8 h or more (lower panel of Fig. 2). If I stood less than 8 h, the probability of early awakening was close to its baseline value. If I stood more than 9 h, the probability of early awakening was near zero.

After I learned that, I tried to stand at least 9 h every day (Phase 2). The probability of early awakening during that phase was close to zero (upper panel of Fig. 2). The dose–response function during Phase 2



**Fig. 2.** Effect of standing on early awakening. Early awakening = fell back asleep within 6 h after getting up. Vertical lines show standard errors. Upper panel: between-phase differences. Lower panel: within-phase differences. Standing durations were divided into three categories: (a) fewer than 8.0 h, (b) 8.0–8.8 h, and (c) more than 8.8 h. The x-axis location of each point indicates the median duration in that category.

was similar to the dose–response function during Phase 1 (lower panel of Fig. 2).

This is unlikely to be a placebo effect because (a) it was surprising and (b) before I discovered the effect of breakfast, I'd tested many treatments that turned out to have no effect at all on early awakening. I had hoped and to some extent expected that all of them would work.

I made more discoveries by self-experimentation. Eventually I had ten. I published a long article about them (Roberts, 2004). Its overall point was that self-experimentation was a good source of new ideas. As these three examples (breakfast, faces, and standing) show, the new ideas could be quite surprising and have practical value.

I couldn't have come up with them without self-experimentation. Maybe that's obvious. The breakfast discovery came from ten years of sleep measurement. Doing sleep experiments with other people as subjects is very difficult. To measure someone else's sleep for two consecutive nights would have been very hard. I measured my sleep for thousands of consecutive nights. Much the same could be said about the faces and standing experiments. I have estimated that self-experiments are about a thousand times easier than conventional experiments (Roberts, 2010). The implications of such an improvement are great. Imagine a microscope that is 1000 times more powerful than existing ones. Or imagine a new way to travel that is 1000 times easier than existing methods.

### 3. Reception within my profession

Unfortunately, I expected my colleagues (other psychology professors) to dislike this work. Anyone who isn't a psychology professor is likely to notice two things about it: (a) I studied myself and (b) the practical value of the conclusions. A likely reaction is that the conclusions are fascinating and that time will tell whether self-experimentation is reliable.

My colleagues, I believed, would see things quite differently. The self-experimentation they would shrug off. Within academic psychology, self-experimentation is rare but exists (Ebbinghaus, 1915; Kristofferson,

1976, 1977, 1980; Voss, 2009). The practical value of the work would be ignored. Academic psychology almost never has a clear practical value and research isn't judged on that dimension.

What would matter to psychology professors, I thought, was that the research was outside my area of expertise. They would greatly dislike this. My area of expertise, the area in which I received graduate training and had published many articles, was animal learning. Roberts (2004) contained data and theory in three areas (sleep, mood, and weight) outside animal learning including a new theory of mood and a new theory of weight control. The history of psychology contains nothing like this, not even close. It's like a dog talking. Now and then a researcher might move from one area to a neighboring one – from animal to human learning, for example. That's fine. Moving to a non-adjacent area, as I did – sleep, mood, and weight control are not close to animal learning – is very rare, partly because it would be hard to have a career (graduate students, grants) in the new area. I knew of only two examples. Richard Herrnstein, originally in animal learning, did work on IQ (Herrnstein, 1973; Herrnstein and Murray, 1994). Leon Kamin, also in animal learning – and reacting to Herrnstein's work – also did work on IQ (Kamin, 1974). Both examples were quite different than mine. Herrnstein's work was aimed more at the public than other scientists; Kamin's work was more forensic than scientific. In both cases, the conclusions weren't mainstream psychology. In my case, the conclusions were mainstream psychology, reached in an unorthodox way. And I'd gone outside my expertise four times (sleep, mood, health, and weight) in one paper. It really was like a dog talking. Apart from the shock value, it had to be disturbing. For a chemist to write an ambitious paper in biology wouldn't sit well with biologists. *How dare he!* the biologists would think. *I've spent my life studying this and he comes in and ...* It would be disturbing only if there were merit in it. So there was mental pressure to see it as worthless. No one ever complained that I was outside my area of expertise. But some of the complaints were extraordinarily inaccurate.

Roberts (2004) appeared in the journal *Behavioral and Brain Sciences* (BBS), which has peer commentary with every article. The comments were mostly positive. One made a point I make here: "Self-experimentation allows experimenters a kind of extended access to the behavior of adult human subjects that is difficult – if not impossible – to achieve otherwise" (Glenn, 2004, p. 264). One negative comment was that "there is a large history of errors from self-inspection and self-experimentation" (Voracek and Fisher, 2004, pp. 273–274). If true, that's very important. But the only example of such errors given by the authors was Freud, whose self-experimentation (with cocaine) was nothing like mine. Nor did the authors make clear what Freud's error was. The failure to provide even one good example from a "large history" was curious. Perhaps there were no good examples. David Booth, one of the world's best food psychologists, commented that my theory of weight control "has been refuted by extensive prior work" (Booth, 2004, p. 262), my data is "anecdotes" (Booth, 2004, p. 263), and my weight loss could have been due to an upset stomach. The upset-stomach explanation didn't make sense because I'd said "I felt fine" (Roberts, 2004, p. 254) during the initial weight loss. My theory of weight control had led me to a powerful and counter-intuitive way of losing weight – a huge success for that theory. No other theory of weight control had ever done such a thing. For Booth to ignore this was quite odd.

At the time Roberts (2004) was published, I was an associate professor at the University of California, Berkeley. Soon after its publication, I was considered for promotion to professor. The BBS paper was my main argument for promotion. It described a great deal of work, was in a prestigious journal, and used an unusual methodology to do something difficult: generate new plausible ideas. It contained surprising and useful new ideas, both methodological and substantive, with experimental support.

The promotion committee (two persons) assigned to my case decided it was worthless. Is self-experimentation a good source of new ideas, as I claimed? No, said the committee.

"In Prof. Roberts' own examples, the self-experimentation followed the generation of the idea. Thus it is difficult to understand how, as claimed, this work contributes to our understanding of the process by which new scientific ideas are generated."

Not quite. In Example 1 (breakfast), data I collected during self-experimentation had a surprising pattern (a correlation between weight and sleep duration) that led to a new idea. In addition, changing what I ate for breakfast had surprising effects, which led to another new idea. In both cases, self-experimentation led to a new idea. In Example 2 (faces and mood), I did a self-experiment hoping for one result (sleep improvement). An unexpected outcome (mood improvement) led to several new ideas about mood. In Example 3 (standing), a self-experimental attempt to lose weight by standing more plus my sleep records (which existed due to self-experimentation) led me to discover that standing reduced early awakening – a new idea. In Example 4 (morning light), I tried getting more morning light to see if it would raise my mood. To my surprise, it improved my sleep – a new idea. In Example 5 (health), my self-experimental records led me to conclude that better sleep had improved my health – a new idea.

In Examples 1–5, then, the pattern *self-experiment* → *new idea* was everywhere. In most cases, the new idea led to further self-experimentation, so that the overall pattern was *self-experiment* → *new idea* → *self-experiment*. Examples 6–10 (weight control) do fit the pattern 1. *Think of a new idea*, 2. *Test via self-experiment* – that is, *new idea* → *self-experiment*. The committee's description ("the self-experimentation followed the generation of the idea") ignored half of the examples.

The statement also ignored, or misunderstood, the point of Examples 6–10. When the title of Roberts (2004) called self-experimentation "a source of new ideas" I meant *plausible* new ideas. *Implausible* new ideas are useless. You can think of a dozen in ten minutes. In Examples 6–10, I used self-experimentation to test predictions that would have otherwise been impossible to test. Two plausible new ideas emerged from those tests: 1. A theory of weight control centered on flavor-calorie associative learning. No previous theory of weight control involved associative learning. 2. A practical new way of losing weight (Example 10, drinking fructose water). To lose weight by eating more of a certain food was quite different from previous methods, which usually involved eliminating certain foods or eating less. Before Example 10, the notion that sugar can cause weight loss was very implausible. Everyone believed sugar causes weight gain. The data of Example 10 made plausible an idea that had been highly implausible.

The committee went on to say that I did not "really grapple with [the] impact [of experimenter bias effects] on self-experimentation as a means of collecting scientific data." In fact, I pointed out four reasons that experimenter bias – which pushes results toward what the experimenter wants or expects – was unlikely to explain my results: (a) Experimenter bias cannot produce surprises. Most of the results, and all of the main ones, were surprising. (b) The main conclusions were supported by results from studies where experimenter bias was unlikely to have been a problem – experiments with rats, for example. (c) In some cases I tried many things that didn't work. For example, I tried many ways to reduce early awakening. When something finally worked, it's highly unlikely that it worked simply because I wanted it to work. I wanted everything I tried to work. (d) When placebo effects have been measured, they have been usually weak or non-existent. The committee ignored these arguments.

The committee also said that "self-experimentation by its very nature cannot address questions of generalizability, external validity, and ecological validity". Generalizability and external validity are the same. My work was *more* ecologically valid (that is, realistic) than most experimental research because it was not done in a lab. Researcher introspection has the same strength because the observations are gathered outside of a lab. Self-experimentation can *better* answer questions about ecological validity than most research because it can more easily be done under realistic conditions. The committee's statement makes no sense. As for generalizability, I had

included evidence from other studies (involving other subjects) that supported my conclusions. For example, my conclusion that breakfast caused early awakening was supported by research with rats. That was good reason to think my breakfast results would generalize to other people. Sure, self-experimentation by its nature cannot explore generalizability over people, but it can explore generalizability over other aspects of the experiment. If I study the effect of butter, for example, I can test different kinds of butter.

The committee went on to criticize my weight-control ideas: "Eating bad-tasting food may lead to sustained weight loss, but only in those who have the self-discipline to adhere to such a regimen." My conclusions were not likely to generalize, in other words, because hardly anyone will have enough self-discipline. This was highly misleading. None of my examples involved bad-tasting food. And my ideas were a lot less obvious than "eating bad-tasting food may lead to sustained weight loss." I'd proposed and tested a theory of weight control centered on associative learning, a big departure from previous theories. I showed that the theory was supported by my results, by evidence from other humans, and by evidence from rats. It had suggested new ways of losing weight that turned out to work. The committee ignored this.

The committee added a final reason not to promote me. The impact of this work was uncertain, they said, and therefore the department "should delay action on [that is, deny] Prof. Roberts' promotion request until we can determine what the field's reaction to his work is." I know of no other case where research in a prestigious journal – or any journal – was given zero value until "the field's reaction" was known.

The promotion committee's negative recommendation was upheld by the other professors in the department 22 to 1. I didn't even get a *within-rank* promotion (from one level within associate professor to a higher level). You get a within-rank promotion if you publish a few average-length articles in average journals.

Low opinion of my work wasn't limited to my Berkeley colleagues. In 2007, at a convention of experimental psychologists, I gave a talk about later self-experimentation in which I'd studied the effect of omega-3 on brain function. I described two experiments and four measures of brain function. During the question period, one person angrily called my talk "pop psychology," criticized it in a few ways, and said I shouldn't have been allowed to speak. *Which objection would you like me to answer first?* I asked. *I don't care what you say*, he said, and walked out. I've seen hundreds of scientific talks. I've never seen anything like that. At a reception the next day, a woman approached me. She'd been at my talk. *Did you control for learning?* she asked. Of course I had. I'd done learning experiments for thirty years. Maybe she was trying to start a conversation.

#### 4. Reception outside my profession

Outside my profession, I assumed no one would notice my BBS paper. However, Andrew Gelman, a friend of mine who's a statistics professor at Columbia, blogged about it shortly after publication. He wrote:

I liked the paper... I was really impressed, first by all the data (over 50 (that's right, 50) scatterplots of different data he had gathered), and second by the discussion and interpretation of his findings in the context of the literature in psychology, biology, and medicine.

Alex Tabarrok, an economics professor at George Mason University, read this. A few weeks later, he wrote about my BBS paper on his popular blog *Marginal Revolution*. The title was "Seth Roberts is Utterly Mad (but in a good way)". He wrote:

A virtue of self-experimentation is that it doesn't take a million dollar lab and a bevy of graduate students, with some willpower and a willingness to carefully document and measure results, anyone can do cutting-edge science.

Which is a long way from the promotion committee's view that my paper "contributes [nothing] to our understanding of the process by which new scientific ideas are generated."

Thousands of people read Tabarrok's post. Among them were Stephen Dubner and Steven Levitt, the authors of *Freakonomics*. They wrote a favorable article about my work in their *New York Times Magazine* column (Dubner and Levitt, 2005). After it appeared, several book agents called me. They believed they could sell a weight-loss book based on my ideas. I soon got a book contract and wrote *The Shangri-La Diet* (Roberts, 2006). It was a *New York Times* best-seller for one week and has been translated into eight languages. About 100,000 copies have been sold.

Publication of a popular book made other things possible. I started a forum (boards.shangriladiet.com) connected with the book, mainly for Shangri-La dieters. After four years, it has almost 100,000 posts from 5000 members and gets 300,000 hits/month. The stories in the forum make it clear that the diet, strange as it may sound, works for many people. Countless posts say how surprised and pleased the dieters were.

I started a blog (blog.sethroberts.net), where I write about self-experimentation, scientific method, the Shangri-La Diet, and related topics. It gets about 40,000 visitors per month. It supplies a sense of being listened to, helps develop ideas bit by bit, and provides feedback. I can post self-experimental results and learn what happens when other people try the same thing. For example, after I discovered that flaxseed oil improved my balance, I found out what happened when several other people tried drinking it.

My BBS paper has been downloaded about 30,000 times. For several years, until rankings became unavailable, it was usually the second or third most downloaded paper (in terms of average downloads/week since posting) in the University of California electronic repository.

The success of *The Shangri-La Diet* made me decide (relatively young) to become a professor emeritus at Berkeley. This meant no teaching, no graduate students, and less salary. It would be harder to do empirical research, but easier to write more books.

## 5. Professor in China

It turned out to be lonely sitting at home writing another book (about self-experimentation). A year after I became emeritus, Kaiping Peng, a Berkeley colleague, offered me a job in the to-be-reestablished psychology department of Tsinghua University, the top university in China. (The department had been closed around 1950 to make Tsinghua more specialized.) In 2004, I had taught at Peking University, near Tsinghua, so I had some idea what I was getting into.

The Tsinghua undergraduates were one attraction. In China, high-school students who want to attend college take a national exam. Admission to Tsinghua requires a score in the top 0.01% (1 in 10,000), which corresponds to an IQ of 155. Genius-level IQ is considered 150 or more. From 2004 to 2006, more Tsinghua undergraduates got American Ph.D.s than undergraduates from any other school, foreign or American (Mervis, 2008). Compared to Berkeley, I could expect more of the undergrads I taught to become psychology professors and do research. I would only have to teach one semester per year, which allowed time for book writing.

Tsinghua graduate students would be much worse than Berkeley graduate students. But this didn't matter because at Berkeley, it was hard for me to have graduate students. No one could make a career out of self-experimentation. Because I had no reputation in the areas I wanted to study, such as sleep and weight control, doing research about them with me wouldn't help you get a job (in America). At Tsinghua, however, three things would help my graduate students get jobs: psychology's growing popularity in China, the great expansion of Chinese higher education, and Tsinghua's prestige.

Teaching brilliant (or at least highly academic) undergraduates has been a pleasure. Last year, I assigned my BBS paper in a freshman

seminar and asked the students to comment on it. A student named Jiang Zhaomeng wrote this:

As a freshman I learn a lot from the examples, not only the facts and results, but also something like spirits.

First, its creativity. Some of the thoughts and results are not mentioned in those authoritative writings and theories, but the author discovered them and had the courage to test them. Maybe they can't be agreed by all the people, even critical words are more than appreciative ones at first, but noticing, discovering and announcing them is already the success of creativity and courage, which are exactly my shortness, I think.

Second, it's the method of experiment ourselves. I think it need enthusiasm and devoted contributions to the science. It's the key which we need to study further. Self-experimentation has many advantages. We can obviously and directly know about the effects and results. We can also adjust the method to our own conditions during the experiments.

Third, I learn that we can get inspiration from our daily life. Like the Example 6, 9 and 10, the author gets ideas from his friend, meal and travel experience. We can also discover particular phenomenon and get ideas from our daily activities. Newton got the law of universal gravitation from the fallen apple. Science comes from life. It's true and makes science interesting.

She made the case for my work better than I have.

## 6. Like-minded groups

None of this, except Jiang Zhaomeng's comments, may surprise you. It's not the first time something new has been dismissed by insiders, only to find an audience elsewhere. But two other developments – the establishment of two groups sympathetic to my work – fit no obvious pattern.

In 2008, Kevin Kelly and Gary Wolf, both associated with *Wired* magazine, started a San Francisco Meetup group called The Quantified Self ("tools for knowing your own mind and body," says the website). It meets every two months. The meetings consist mostly of short talks about high-tech self-measurement. One recent speaker recorded how many Unix commands he wrote at various times of day; another used an iPhone to track his meditation time; a third used a heart-rate monitor to show himself how nervous he was. The idea spread quickly. By 2010, seven other cities (London, New York, Chicago, Amsterdam, Boston, Seattle, and Sydney) had Quantified Self groups. It has grown into a tiny organization, with an employee and board of directors. The long-term prospects are good. Blood-sugar tracking has greatly helped diabetics. I am not diabetic, but blood-sugar measurements helped me discover how to improve my blood sugar levels (by walking an hour/day – see my blog for details). It's plausible that tracking this and other health measures, such as blood pressure and sleep quality, could greatly benefit many people. The Quantified Self may be at the beginning of this.

Equally promising is the 2010 founding of the Ancestral Health Society, which will encourage an evolutionary approach to health. My self-experimentation played part in this. As I said earlier, it led me to believe that solutions to many health problems may be found among the elements of Stone-Age life, such as no breakfast. Part of Roberts (2004) is titled "Stone-Age Life Suits Us".

I wrote about this on my blog. Aaron Blaisdell, a professor of psychology at UCLA and a collaborator of mine (Stahlman et al., 2010), read my blog and became interested in this point of view. He discovered Weston Price's work, the Weston Price Foundation, and the paleo/primal communities. He tried their recommendations, such as eating more cod-liver oil, high-vitamin butter oil, and fermented dairy (e.g., yogurt), and eating less grains (especially gluten), legumes (peanuts and beans), and industrial oils (such as corn oil). He was

deeply impressed how much better he felt. To his surprise, his porphyria went away. He had had erythropoietic protoporphyria, a genetic condition that meant he couldn't spend more than 5 or 10 min in the sun. After he made the nutritional changes, the problem disappeared. He could spend an hour in the sun.

Aaron and Brent Pottenger, a recent college graduate planning to be a doctor, decided it would be good to have a symposium about this approach. Aaron, Brent, and I have been organizing such a symposium (to be held in 2011). A desire to have a sponsor for the symposium and a big response to its announcement led Aaron, Brent, and others, including me, to found the Ancestral Health Society, which will hold future symposia and promote these ideas in other ways. It resembles a professional society but has attracted many volunteers. In the summer of 2010, every few weeks Aaron got an email from someone who had heard about it and wanted to help.

Both groups provide a scientist's most basic need: an audience. Maybe they will help me find collaborators. By making my work part of a larger movement, they will help it be more influential.

## 7. What I learned

I came to see my work and its reception as examples of historical patterns. One of those patterns is *the innovative power of insider/outside*s. An insider/outside is someone with the knowledge of an insider and the freedom of an outsider. I was an insider/outside with respect to the topics of my research (sleep, etc.). I had insider-like knowledge because I was a professor of psychology. Although I wasn't a sleep, mood, or weight expert, I taught about sleep, mood, and weight in my classes. I was skilled at experimental design and data analysis. I had the freedom of an outsider in the sense that I could do whatever self-experiments I wanted. Professional researchers are much more constrained than I was (Roberts, 2010). To get another grant, to appear productive, to attract graduate students, maybe even to keep their job, they must publish regularly. A less obvious constraint is that, like everyone else, they want to appear high-status. Veblen (1899) argued that doing expensive useless research marks you as higher status than doing cheap useful research. I believe it is hard for professional researchers to do cheap useful self-experimentation, such as mine, because they think it would make them look low-status. Researcher introspection may or may not have a clear practical value but it is inevitably low cost, which creates a similar problem.

Science owes a lot to insider/outside. Charles Darwin had insider-like knowledge because he'd had an excellent education in biology. He had the freedom of an outsider because he was wealthy and jobless. He could safely propose radical ideas. Gregor Mendel had scientific training. His colleagues (monks) didn't care what he wrote about plants. Alfred Wegener, who proposed the theory of continental drift, was a meteorologist.

I came to believe that it was the *combination* of self-experimentation and my insider/outside status that was powerful. Anyone can self-experiment. They're unlikely to make important discoveries because, unlike insiders, they can't choose wisely what to study and can't do a good job of data analysis. Although insider/outside are in a good position to innovate, they rarely have the necessary resources – because they're outsiders. I had an insider-like knowledge of sleep research but, being an outsider, could never get a grant to do research. Self-experimentation empowered me.

Is science better as a hobby or a job? Well, you need great freedom, especially freedom to fail, to be a good scientist. Hobbyists have that freedom, professionals do not. Yet you also need considerable technical knowledge to be a good scientist. Which you presumably need to work full-time to acquire. So good science is an accident, an anomaly.

Why do outsiders do it? Darwin, Mendel, and Wegener didn't reap professional rewards from their work. They weren't paid. They weren't promoted. They didn't win awards or recognition. What motivated them?

Before jobs, there were hobbies (Roberts, 2005). I suspect Darwin, Mendel, and Wegener were motivated by the same preferences and desires that drove the first hobbyists, hundreds of thousands of years ago. An example of those motivations is Veblen's "instinct of workmanship" (Veblen, 1914, p. 1) – a desire to make well-made things quite apart from how much you can sell them for. I've proposed that humans evolved from other primates in two steps (Roberts, 2005). First came an age of hobbies. Using their manual dexterity, our hunter-gatherer ancestors got better and better at making things. It was a hobby-like activity. Then trading began. Now hobbyists could specialize. They traded their specialty for the specialties of others. As trading and technical knowledge grew, hobbies became part-time jobs and later full-time jobs – bringing us to the current age of occupational specialization, with thousands of ways to make a living. The move to specialization changed our brains but our hobbyist instincts – and their innovative force – remained.

What Darwin, Mendel, and Wegener did had a lot in common with hobbies. Mendel's pea plants looked exactly like a hobby. Darwin's and Mendel's research required many years of work with no visible output. Few jobs allow so little tangible progress, but this is perfectly acceptable for hobbies. Like hobbyists, they were under no pressure to publish, no pressure to conform. They had no fear of losing their job or their colleagues' respect. They had hobbyist freedom.

Darwin, Mendel, and Wegener had insider knowledge, outsider freedom, and hobbyist motivation. I had insider knowledge, outsider freedom, and something more powerful than hobbyist motivation: My self-experimentation improved my health. I slept better, was in a better mood, got sick less often, and lost weight. I published my work in a mainstream journal. It looks like other papers: it has data, graphs, literature reviews, and so on. But it arose from a situation with two unique features: I had outsider freedom; and my work improved my health. I don't know of any other paper in psychology like that. These features, like powerful tools only I could use, allowed me to build something unlike anything built before. My work surely struck my colleagues as exceedingly strange. I think this is another reason, in addition to the fact that it was outside my research area, that it received a poor reception.

If I'm going to innovate like this – with insider knowledge, outsider freedom, and motivated by improvements in health – it makes sense to write books. Books are the natural home for insider/outside expression. Book writing is almost always a hobby, in the sense that almost all authors make a pittance and have a different full-time job. (An exception is that the authors of genre books, such as romance novels, can support themselves by book-writing. Genre books can be written quickly and have a predictable market.) Because payment is insignificant, authors can say almost anything. They have nothing to lose.

Books, unlike any other important piece of our economy, aren't made for profit. Leaving aside genre books, most books are non-fiction; most are based on what the author does for a living. Textbooks about X are written by people paid to teach X. Journalists write books on topics they were paid to cover. Doctors write about medicine. The combination of expertise and freedom, plus distribution via the book industry, makes books highly influential. Darwin and Wegener wrote books. They gained attention right away. Mendel did not write a book. For a long time no one noticed his work.

Another historical pattern is *new ways to spread ideas help innovators*. New ideas need converts. New ways of spreading an idea allow it to reach more people, making it easier to find converts. This is especially true when the new idea is unpopular with those in power. New methods of dissemination will broadcast the idea further from Place X, where the influence of those in power at X will be weaker. I doubt it's a coincidence that the Protestant Reformation began (1517) soon after Gutenberg invented the printing press (1450). That Martin Luther could print his protest (95 Theses) made it much easier to spread. Likewise, the American Civil Rights movement made great progress in the 1950s and 60s, when television became

popular. Protests in the South made the evening news. TV viewers in the North, who didn't care what the Southern establishment thought of them, became a powerful force for change. In my case, the Internet, open-source publishing, and blogging allowed my work to reach people outside my profession.

A final historical pattern is *travel helps innovators*. Like new broadcast media, it helps them find an audience. After it became possible to cross the Atlantic, Englishmen came to New England for religious freedom – freedom to worship new religions. In the 1940–50s, Edward Deming, an American statistician, had new ideas about how to improve manufacturing. In America, he was ignored. In Japan, he was taken seriously. Japanese car makers used his ideas long before American car makers did (Gabor, 1990). My position at Tsinghua and Jiang Zhaoming's comment suggest that I will have an easier time getting my ideas taken seriously in China than in America.

Professors tend to aim for professional recognition, such as winning a Nobel Prize. One day in Paris, I got a glimpse of something different. I was walking down a street when I heard "Stop, thief!" (in French). A man ran towards me. It had just rained and the cobblestones were wet. Trying to get in front of him, I slipped and fell. He fell over me – a perfect tackle. "Why did you do that?" he asked. He got up and ran off. A woman came up: "Are you okay?" As I walked away, a man said, "That was an unusual thing you did. Most people wouldn't have done that." My ineffectual good deed had been noticed. For the next few hours, I was surprised how good I felt. Good deed *plus* praise is a potent combination, I realized. More powerful than good deed alone or praise alone.

In academia, it's hard to be helpful *and* praised. Good teaching (helpful) gets little praise. Most Nobel Prizes (praise) go to work with little practical value (Cassidy, 2009). The scientists who discovered that smoking causes lung cancer (helpful) never got a Nobel Prize. Unexpectedly, I managed both. I helped people lose weight *and* felt recognized. Recognition came from the Shangri-La Diet forums, book sales, blog readership, email from strangers, Jiang Zhaoming's comments, and audiences at Quantified Self meetings. In the future, it may come from the Ancestral Health Society. This is an ecosystem that didn't exist in 2004 (e.g., my blog didn't exist).

A final lesson was *the value of being unsure*. As I wrote Roberts (2004), I believed it would be ignored. But I knew my beliefs were often wrong, so I kept writing.

## References

- Booth DA. How observations on oneself can be scientific. *Behav Brain Sci* 2004;72:262–3.
- Cassidy J. How markets fail: the logic of economic calamities. New York: Farrar, Strauss & Giroux; 2009.
- Dubner SJ, Levitt SD. Does the truth lie within? *New York Times Magazine* [September, 11, 2005, 20–2].
- Earl PE. Simon's travel theorem and the demand for live music. *J Econ Psychol* 2001;22:335–8.
- Ebbinghaus H. *Memory: a contribution to experimental psychology*. New York: Teachers College, Columbia University; 1915.
- Gabor A. *The man who discovered quality: How W. Edwards Deming brought the quality revolution to America: the stories of Ford, Xerox, and GM*. New York: Random House; 1990.
- Glenn SS. Linking self-experimentation to past and future science: extended measures, individual subjects, and the power of graphical presentation. *Behav Brain Sci* 2004;27:264.
- Gould S. The self-manipulation of my pervasive, perceived vital energy through product use: an introspective-praxis perspective. *J Consum Res* 1991;18:194–207.
- Halmos P. The problem of learning to teach: the teaching of problem solving. *Amer Math Mon* 1975;82:466–70.
- Herrnstein R. *I.Q. in the meritocracy*. New York: Little, Brown; 1973.
- Herrnstein R, Murray C. *The bell curve: intelligence and class structure in American life*. New York: Free Press; 1994.
- Hirschman EC. The day I almost died: a consumer researcher learns some lessons from a traumatic experience. In: Hirschman EC, editor. *Research in consumer behavior*, Vol. 4. Greenwich, CO: JAI Press; 1990. p. 115–28.
- Kamin LJ. *The science and politics of IQ*. New York: Lawrence Erlbaum; 1974.
- Kristofferson AB. Low variance stimulus-response latencies: deterministic internal delays? *Percept Psychophys* 1976;20:89–100.
- Kristofferson AB. A real-time criterion theory of duration discrimination. *Percept Psychophys* 1977;21:105–17.
- Kristofferson AB. A quantal step function in duration discrimination. *Percept Psychophys* 1980;27:300–6.
- Mervis JJ. U.S. graduate training: Top Ph. D. feeder schools are now Chinese. 11 July. *Science* 2008;321:185.
- McKibben W. Comment. *The New Yorker*; 1983. p. 43–4. December 12.
- Reid R, Brown S. I hate shopping! An introspective perspective. *Int J Retail Distribution Management* 1996;24(4):4–16.
- Roberts S. Self-experimentation as a source of new ideas: ten examples about sleep, mood, health, and weight. *Behav Brain Sci* 2004;27:227–88. <http://scholarship.org/uc/item/2xc2h866>.
- Roberts S. Diversity in learning. *Ideas That Matter* 2005;3(3):39–43. [http://www.sethroberts.net/about/2005\\_diversityinlearning.pdf](http://www.sethroberts.net/about/2005_diversityinlearning.pdf).
- Roberts S. *The Shangri-La diet*. New York: Putnam; 2006.
- Roberts S. The unreasonable effectiveness of my self-experimentation. *Med Hypoth* 2010;75:482–9.
- Stahlman WD, Roberts S, Blaisdell AP. Effect of reward probability on spatial and temporal variation. *J Exp Psychol Anim Behav Proc* 2010;36:77–91.
- Veblen T. *The theory of the leisure class: an economic theory of institutions*. New York: Macmillan; 1899.
- Veblen T. *The instinct of workmanship and the state of the industrial arts*. New York: Macmillan; 1914.
- Voracek M, Fisher ML. The birth of a confounded idea: the joys and pitfalls of self-experimentation. *Behav Brain Sci* 2004;27:273–4.
- Voss JL. Long-term associative memory in man. *Psychon Bull Rev* 2009;16:1076–81.
- Wallendorf M, Brucks M. Introspection in consumer research: implementation and implications. *J Consum Res* 1993;20:339–59.