BEHAVIORAL AND BRAIN SCIENCES (2004) **27,** 227–288
Printed in the United States of America

Self-experimentation as a source of new ideas: Ten examples about sleep, mood, health, and weight

Seth Roberts
Department of Psychology, University of California, Berkeley, Berkeley, CA 94720-1650.
roberts@socrates.berkeley.edu

Abstract: Little is known about how to generate plausible new scientific ideas. So it is noteworthy that 12 years of self-experimentation led to the discovery of several surprising cause-effect relationships and suggested a new theory of weight control, an unusually high rate of new ideas. The cause-effect relationships were: (1) Seeing faces in the morning on television decreased mood in the evening (>10 hrs later) and improved mood the next day (>24 hrs later), yet had no detectable effect before that (0–10 hrs later). The effect was strongest if the faces were life-sized and at a conversational distance. Travel across time zones reduced the effect for a few weeks. (2) Standing 8 hours per day reduced early awakening and made sleep more restorative, even though more standing was associated with less sleep. (3) Morning light (1 hr/day) reduced early awakening and made sleep more restorative. (4) Breakfast increased early awakening. (5) Standing and morning light together eliminated colds (upper respiratory tract infections) for more than 5 years. (6) Drinking lots of water, eating low-glycemic-index foods, and eating sushi each caused a modest weight loss. (7) Drinking unflavored fructose water caused a large weight loss that has lasted more than 1 year. While losing weight, hunger was much less than usual. Unflavored sucrose water had a similar effect. The new theory of weight control, which helped discover this effect, assumes that flavors associated with calories raise the body-fat set point: The stronger the association, the greater the increase. Between meals the set point declines. Self-experimentation lasting months or years seems to be a good way to generate plausible new ideas.

Keywords: breakfast; circadian; colds; depression; discovery; fructose; innovation; insomnia; light; obesity; sitting; standing; sugar

Mollie: There has to be a beginning for everything, hasn't there?
–*The Mousetrap,* Agatha Christie (1978)

1. Introduction

1.1. Missing methods

Scientists sometimes forget about idea generation. "How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service," wrote Charles Darwin to a friend (Medawar 1969, p. 11). But where did the first views come from, if not observation? According to a diagram in the excellent textbook *Statistics for Experimenters* (Box et al. 1978), the components of "data generation and data analysis in scientific investigation" (p. 4) are "deduction," "design," "new data," and so on. Scientific investigation, the diagram seems to say, begins when the scientist has a hypothesis worth testing. The book says nothing about how to obtain such a hypothesis.

It is not easy to come up with new ideas worth testing, nor is it clear how to do so. Table 1 classifies scientific methods by goal (generate ideas or test them) and time of application (before and during data collection or afterwards). The amount written about idea generation is a small fraction of the amount written about idea testing (McGuire 1997), and the amount written about what to do before and during data collection is a small fraction of the amount written about what to do afterwards – so the empty cell in Table 1, on how to collect data that generate ideas, is no surprise. Although scientific creativity has been extensively studied (e.g., Klahr & Simon 1999; Simonton 1999), this research has not yet suggested new tools or methods. Even McGuire, who listed 49 "heuristics" (p. 1) for hypothesis generation, had little to say about data gathering.

Hyman (1964) believed that "we really do not know enough about getting ideas even to speculate wisely about how to encourage fruitful research" (p. 28), but 40 years later this is not entirely true. Exploratory data analysis (Tukey 1977) helps reveal unexpected structure in data, and such discoveries often suggest new ideas worth studying. Table 1 includes only those methods useful in many areas of science, omitting methods with limited applicability

SETH ROBERTS is an Associate Professor in the Department of Psychology at the University of California at Berkeley. He is a member of the University's Center for Weight and Health. His research includes follow-up of the results reported in this article, especially the weight and mood results, and study of how things begin in another situation – the generation of new instrumental behavior by rats and pigeons.

 0140-525X/04 $12.50

Table 1. *Missing methods*

	Time period	
Goal	Before and during data collection	After data collection
Generate ideas	?	Exploratory data analysis
Test ideas	Experimental design	Statistics (e.g., *t* test)

(e.g., clinical trials). If domain-specific methods were included, Table 1's empty cell would have at least one entry: combinatorial chemistry, which "came of age" in the 1990s (Moos 1998, p. 6). Combinatorial chemistry consists of methods to rapidly make large numbers of chemicals with new structures, which are then screened for new catalysts, materials, sensors, and drugs. This article argues that long-term self-experimentation – lasting months or years – is another instance of the missing type of method: a way to gather data to generate ideas.

1.2. Long-term self-experimentation

Participation in one's own experiments is common in human experimental psychology and usually goes unremarked when other subjects are also tested. It is when the experimenter does something that would have been hard to do with other persons that attention is paid and the term *self-experimentation* is applied (e.g., Altman 1987/1998). That is how the term will be used here.

Long-term self-experimentation, in this sense, is rare. Bouts of self-experimentation have usually been brief (Altman 1987/1998; Carpenter et al. 1997). For example, when Barry Marshall, an Australian gastroenterologist, drank *Helicopter pylori* bacteria to show that they can cause ulcers (Monmaney 1993), the experiment took only about two weeks to confirm it (Sobel 2001). Nevertheless, the self-experimentation work described in this target article, which lasted 12 years, had several precursors:

1. Santorio Santorio (born 1561), an Italian doctor, discovered insensible perspiration by measuring his intake and excretions over 30 years (Altman 1987/1998).

2. Ebbinghaus (1885/1913) began the experimental study of memory with work done over a 5-year period.

3. In 1961, Victor Herbert, an American medical researcher, ate a folate-free diet for 5 months to determine the effects of folate deficiency (Brown 1995; Herbert 1962).

4. In 1962, Michael Siffre, a French cave explorer, lived in a cave for 2 months without time-of-day information, measuring his sleep and other variables (Siffre 1964).

5. Kristofferson's (1976; 1977; 1980) duration discrimination work lasted more than 3 years, judging by manuscript submission dates.

6. Franz Halberg and other chronobiologists have self-measured blood pressure, heart rate, and other physiological variables at closely spaced intervals for long periods of time, often many years (e.g., Cornélissen et al. 1997; 1999; Halberg et al. 1965; 2000).

7. Blood glucose self-monitoring by diabetics, now common, originated with Richard Bernstein, an American engineer with insulin-dependent diabetes (Bernstein 1997). Four years of trial and error led him to insulin and diet regimes that kept his blood glucose level much more constant than the standard advice.

The present work resembles these precursors in several ways, including the use of long-term records (Santorio, Halberg et al.), unusual diets (Herbert, R. Bernstein), measurement of biological rhythms (Siffre, Halberg et al.), and practicality (Herbert, Halberg et al., R. Bernstein). Unlike any of them, the present work relied heavily on anecdotes and accidents (unexpected outcomes) as starting points for research. Accidents have often been the starting point for conventional research, of course (e.g., Beveridge 1957; Roberts 1989; Valenstein 1998).

In spite of Ebbinghaus, psychologists currently have a mild taboo against self-experimentation (e.g., Baer 1998). One reason is concern that expectations may influence results (Baer 1998; Rosenthal 1966). The use of accidents as starting points is helpful here because, by their nature, accidents cannot be due to expectations. Tests to confirm an accident may be influenced by expectations, however, so the problem is not entirely solved. In any case, the taboo against self-experimentation makes it harder to learn what it is good for. I argue here that a great strength of self-experimentation – its ability to generate plausible new ideas – has been overlooked.

1.3. Background and overview

I began self-experimentation in graduate school to learn how to do experiments (Peden & Keniston 1990; Roberts 2001). Unexpectedly, some of what I did was useful, helping to reduce my acne (Roberts 2001; Roberts & Neuringer 1998). After that, I thought of self-experimentation as advanced trial and error that might solve everyday problems. I spent years testing changes that I thought might reduce how often I woke up too early. The results (almost all negative) seemed unlikely to interest anyone else. But then I found that breakfast, morning faces, and standing (see Examples 1–3) had surprising effects on sleep and mood, and I realized that the rate of idea generation was remarkably high. More thought-provoking results followed (Examples 4–10). I seemed to have stumbled on a way to gather data to generate plausible new ideas.

In addition to generating new ideas, self-experimentation also pushed them toward plausibility. The general pattern was that self-experimentation and self-observation, lasting years, now and then produced a surprise that suggested a new idea. Brief self-experiments tested the new idea, which became more plausible when confirmed by the tests. In some cases self-experiments added detail to the idea – for instance, by determining what controlled the size of an effect. The added detail helped show how the new idea made sense in terms of older, well-accepted ideas, further increasing its plausibility.

The 10 examples are divided into two sections: "Stone Age life suits us" (Examples 1–5), about the damage done by differences between modern life and Stone Age life; and "Pavlovian weight control" (Examples 6–10), about how food controls weight. Each example has the same structure: introduction, birth of the idea, test or tests (except that Example 5 has no tests), related results, and discussion.

2. Stone-Age life suits us

2.1. Introduction

Around 1980, I began to wake up too early. I sometimes awoke at 4 a.m., still tired but unable to fall back asleep. Only after a few dreary hours would I manage to fall back asleep. This happened several times per week. The usual name for this is *early awakening;* but the trouble was not when I awoke but how tired I felt at the time and how long it took to fall back asleep. Had I had awakened feeling rested or fallen back asleep quickly, waking up early would not have bothered me.

Conventional treatments for early awakening included improving "sleep hygiene" (Vgontzas & Kales 1999, p. 390; e.g., don't drink too much alcohol), psychotherapy, and drugs (Vgontzas & Kales). But my sleep hygiene was already good; psychotherapy seemed irrelevant (I was not depressed); and drugs were dangerous. None of the treatments, as far as I could tell, was based on convincing evidence about cause.

Maybe self-experimentation would help. The problem was easy to measure. Some days I awoke so tired that I fell back asleep a few hours later; on other days I awoke more rested and did not fall asleep until the afternoon, when I often took a nap. Because I left home late in the morning, I could always fall back asleep if I wanted to. If I did not fall back asleep within a few hours after waking, I almost never fell back asleep before noon. So a good measure of the severity of the problem was how often I fell back asleep within a few hours of waking up.

Examples 1–5 describe how self-experimentation did help, though not in an obvious way. This section is titled "Stone-Age life suits us" because in each case something resembling Stone-Age living conditions was beneficial.

2.2. Example 1: Breakfast → Early awakening

2.2.1. Background. My early awakening had started close to the beginning of my use of full-spectrum fluorescent lamps near my bed. To mimic sunrise, timers turned these lamps on one by one, producing a step-by-step rise in illumination early in the morning. The goal was to wake up earlier, and the light did indeed have this effect (Roberts 1994). Because my early awakening and my use of the lamps started at about the same time, I suspected that the lamps caused the early awakening.

That something as natural as sunrise could impair sleep seemed unlikely. So I focused on a difference between the bedroom light and sunrise: The lamps went from zero to full brightness in three steps, whereas the intensity of sunlight increases gradually. Maybe the square-wave onsets increased the amplitude of a light-sensitive ultradian rhythm, and this caused early awakening. Some of my data suggested that an oscillator with a cycle length near 90 minutes (e.g., Klein 1979; Orr et al. 1974) influenced when I woke up and fell back asleep.

To test this idea, I randomly varied the onset time of the lights to eliminate any consistent phase relation between light onset and a 90-minute oscillator. To my surprise, this did not eliminate early awakening. Puzzled, I tested the assumption on which my speculation about a 90-minute oscillator had been based: that the bedroom fluorescent light caused early awakening. I had not previously tested this idea because I had been so sure of it. It too turned out to be wrong: Early awakening *increased* when I eliminated the fluorescent light (see Example 4 for details). This was baffling. I had tested other possible solutions, such as exercise and diet (e.g., more calcium). None had worked. With all my ideas about causation proved wrong, I could think of nothing else to do. I was in exactly the situation that conventional books about scientific method (such as Box et al. 1978) do not address: I had no hypotheses to test.

2.2.2. Birth of the idea. While recording early awakening, I also wrote down other information about my sleep, including when I fell asleep and got up (see the Appendix for details). In 1990, I got a home computer, which made analysis of this data much easier. In 1993, I looked at a graph of sleep duration over time (the upper panel of Fig. 1 is an updated version). It showed that several months earlier my sleep duration had decreased about 40 minutes per day – a change I hadn't noticed.

The decrease had happened at the same time that I had lost about 5 kg by eating less highly-processed food. That is, I ate food that was closer to its natural state (e.g., more fruit, oranges instead of orange juice, rice instead of pasta, vegetables instead of bread, other baked goods, and delicatessen food). Letting Day 1 be the first day of the dietary change, comparison of Days −500–0 to Days 51–550 shows that the decrease in sleep duration was highly reliable, $t(880) = 6.89, p < 0.001$. The lower panel of Figure 1 shows the weight change. The weight measurements ended because the scale broke. I never regained the lost weight.

I showed a graph similar to the upper panel of Figure 1 to my introductory psychology class. It inspired a student to tell me that he had reduced how long *he* slept by eating food high in water content (Michael Lee, personal communication, March 1993). In practice, this meant eating lots of fruit. I tried eating four pieces of fruit each day. My sleep duration did not change. When I told Michael the results, he said, "I eat six pieces of fruit." So I started eating six pieces of fruit each day. To increase my fruit intake, I replaced my usual breakfast (a bowl of oatmeal) with two pieces of fruit, often a banana and an apple or a pear and an apple.

Again, my sleep duration stayed roughly the same. But the change of breakfast had a clear effect: Early awakening *increased.* During the three weeks before I started eating fruit breakfasts, the sequence had been 010010000101101011000, where 1 = a day with early awakening and 0 = a day without early awakening. (By *early awakening* I mean falling back asleep within 6 hours after getting up, a limit that avoided inclusion of afternoon naps.) When fruit breakfasts began (on Sunday), the sequence of early awakening became 00111111111 (the sequence begins on Monday, the morning after the first fruit breakfast). Because I was recording sleep and breakfast on the same piece of paper, the breakfast/early awakening correlation was easy to notice. I switched back to oatmeal for four days: 0010. I tried fruit again for four days: 1111. Apparently a fruit breakfast increased early awakening.

The breakfast/early awakening connection surprised me because I was not hungry when I awoke too early. But it was not mysterious. Food-anticipatory activity is a well-established effect in animals (see sect. 2.2.4, below) and the details of the effect corresponded roughly to what I observed.

The fruit effect was in the wrong direction: I wanted to *decrease* early awakening. Oatmeal and fruit are both high

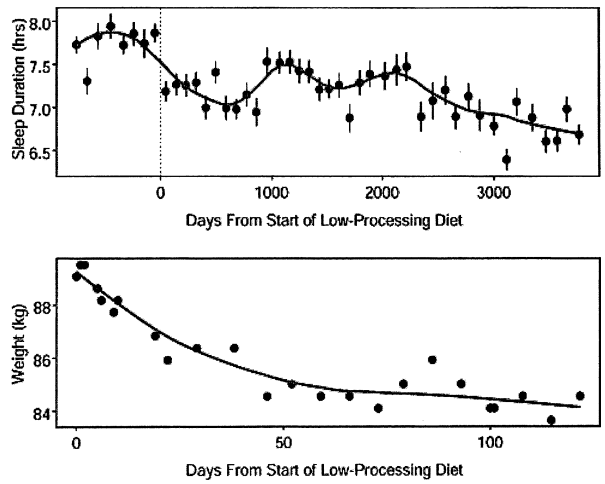

Figure 1. Sleep duration over 12 years (upper panel) and weight loss over 4 months (lower panel). The low-processing diet started March 27, 1992. Each point in the upper panel is a 10% trimmed mean over 84 days. The error bars are jackknife-derived standard errors. Sleep duration includes both sleep at night and naps during the day; it does not include days when I was away from home or sick.

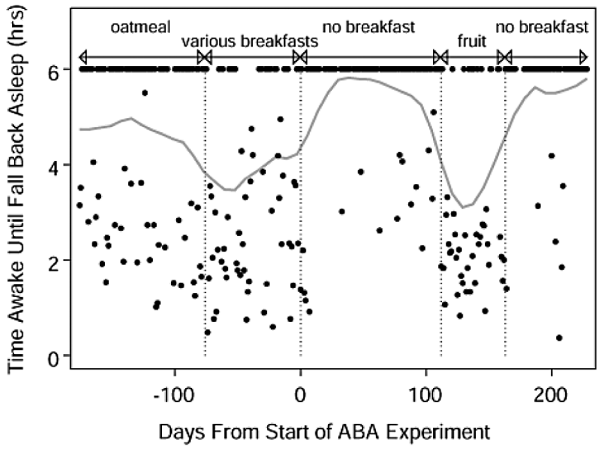

Figure 2. Early awakening over 1 year. The ABA experiment began October 2, 1993. Each point with a y value of less than 6 hours indicates an occurrence of early awakening; days with no early awakening are given a value of 6 hours. Greater average values indicate less early awakening. The line is a loess fit.

in carbohydrate and low in fat, but oatmeal has much more protein than fruit. If less protein was worse, more might be better. I tried several high-protein breakfasts (including shrimp, low-fat yogurt, and hot dogs). None eliminated early awakening, or even seemed to reduce it.

Around this time, I learned of rat experiments that showed that each source of calories (carbohydrate, protein, and fat) can by itself cause anticipatory activity (Mistlberger et al. 1990). This meant I should consider more than protein in my search for a better breakfast.

Food varies on many dimensions (calories, fat, cholesterol, sugar, etc.). I had no idea which mattered. My little experiments had compared something (fruit, high-protein breakfasts) to something else (oatmeal). Perhaps the results would be easier to interpret if I compared something to nothing (no breakfast). So I stopped eating breakfast – had nothing to eat or drink before 11 a.m. – to provide a no-breakfast baseline. To my great surprise, early awakening disappeared. (Or so it seemed at the time. In fact, it just became much rarer.)

Figure 2 shows what happened. During the first no-breakfast phase (Days 1–112 in Fig. 2), the rate of early awakening was (a) reliably lower than during the preceding phase of various breakfasts (Days −75–0), $\chi^2[1] = 36.69$, two-tailed $p < 0.001$, and (b) reliably lower than during the last 100 days of oatmeal breakfasts (Days −175 to −76), $\chi^2[1] = 9.91$, $p < 0.002$.

2.2.3. Test. To confirm that skipping breakfast reduced early awakening, I started eating breakfast again (one piece of fruit between 7 and 8 a.m.). Early awakening increased. Comparing the fruit phase (Days 113–153 in Fig. 2) to the first no-breakfast phase, $\chi^2[1] = 44.80$, $p < 0.001$. I stopped eating breakfast again. Again, early awakening became rare. Comparing the second no-breakfast phase (Days 154–219) to the fruit phase, $\chi^2[1] = 41.00$, $p < 0.001$. Eliminating breakfast clearly reduced early awakening.

2.2.4. Related results. Several studies have linked body fat and sleep duration. After anorexics gained weight, they slept more (Crisp & Stonehill 1971; Lacey et al. 1975). A cross-sectional study found that anorexic women slept less than both normal-weight women and normal-weight women with bulimia, whose sleep durations did not differ (Walsh et al. 1985). After obese patients lost weight, they slept less (Crisp et al. 1973; Ho et al. 1978). After a college student cut her caloric intake in half, she slept less (Neuringer 1981). Crisp and Stonehill (1973) found that psychiatric patients who lost weight slept less than those who gained weight. A survey of Canadian adults found a positive correlation between body fat and hours spent in bed (Shepard et al. 1969).

Food-anticipatory activity is a well-established effect in animals (Bolles & Stokes 1965; Boulos & Terman 1980). Mammals, birds, and fish become more active a few hours before feeding time (Boulos & Terman); as far as I know, no effect present in mammals, birds, and fish has ever been absent in humans. Because activity requires wakefulness, food should produce anticipatory wakefulness as well. Details of what I observed resemble details of the anticipatory-activity effect: (a) *Timing.* Animals become more active a few hours before feeding time (Boulos & Terman). I often ate breakfast around 7 a.m. and often woke up around 4 a.m. (b) *Time course of disappearance.* When I stopped eating breakfast, early awakening went away gradually during the first week of no breakfast. Rat experiments have found that when food stops, anticipatory activity gradually disappears over the next 5 to 10 days (Boulos et al. 1980). (c) *Sufficient foods.* As mentioned earlier, carbohydrate alone and protein alone can produce the effect (Mistlberger et al. 1990). I had found that early awakening could be produced by a high-carbohydrate meal (fruit), a variety of high-protein meals (e.g., shrimp), and a meal with substantial amounts of both carbohydrate and protein (oatmeal).

2.2.5. Discussion. Were the results due to expectations? Surely not. The initial discovery (the effect of replacing oatmeal with fruit) was unexpected, and so were later results. In some cases (the change from oatmeal to fruit, and the change from some breakfast to no breakfast), a change I had expected to have no effect had a large effect. In other cases (the prior failures, the high-protein breakfasts), changes I

had expected to reduce early awakening did not do so. Moreover, the anticipatory-activity results with animals could not be due to expectations.

The main conclusion – skipping breakfast reduces early awakening – was (and is) a new idea in sleep research (Vgontzas & Kales 1999). That it had seemed to come from nowhere was impressive. To test or demonstrate an idea derived from other sources had been a familiar use of self-experimentation (e.g., Marshall drinking *Helicopter pylori* bacteria, as described in Monmaney 1993). To *generate* an idea – in this case, an idea that passed two tests and was supported by other research – had not.

Self-experimentation helped generate the idea in several ways. It made it much easier to collect a long record of sleep duration and early awakening, to notice that the shift in breakfast from oatmeal to fruit coincided with an increase in early awakening, and to eat a wide range of breakfasts.

2.3. Example 2: Morning faces → Better mood

2.3.1. Background. Skipping breakfast did not eliminate early awakening. During the two no-breakfast periods of Figure 2 (October 1993–May 1994), I awoke too early on about 10% of mornings. Over the next few months the rate increased, even though I never ate breakfast (that is, never ate before 10 a.m.). From July to September 1994, I woke up too early on about 60% of mornings. The increase surprised me; I had assumed *a* cause was *the* cause.

What else might cause early awakening? The breakfast effect made me think along new lines. Every tool requires certain inputs, certain surroundings, to work properly. I doubted that our Stone Age ancestors ate breakfast. Before agriculture, I assumed, food was rarely stored. If these assumptions were correct, it made some sense that breakfast interfered with sleep – our brains had been shaped by a world without breakfast.

This sort of reasoning was not new. *Diseases of civilization* (also called *diseases of affluence*) are health problems, such as heart disease and diabetes, with age-adjusted rates higher in rich countries than in poor ones (Donnison 1937; Trowell & Burkitt 1981). The usual explanation is that wealth brings with it a way of life different from the way of life that shaped our genes (Boyden 1970; Cohen 1989; Eaton & Eaton 1999; Neel 1994). For example, "there is now little room for argument with the proposal that health . . . would be substantially improved by a diet and exercise schedule more like that under which we humans evolved" (Neel 1994, p. 355). Such ideas are called *mismatch theories.*

Mismatch theories, however hard to argue with, have serious limitations. First, the speed of evolution is uncertain. The mismatch is between now and when? The answer is often taken to be the late Stone Age, 50,000 to 10,000 years ago (e.g., Eaton & Eaton 1999), but Strassmann and Dunbar (1999) argued for both earlier and later dates. Based on rates of diabetes among Nauru Islanders, Diamond (1992) concluded that human genes can change noticeably in one generation. Second, long-ago living conditions are often unclear. For instance, Cordain et al. (2000) tried to estimate the optimal plant/animal dietary mix by looking at the diets of recent hunter-gatherers, but recent hunter-gatherer diets vary widely (Milton 2000). Third, any prehistoric lifestyle differs from any modern lifestyle in countless ways. Fourth, it is unlikely that long-ago living conditions were ideal. That the average Stone-Age diet contained enough Vitamin C to prevent scurvy seems likely; that it contained the optimum amount of Vitamin C seems unlikely. That it always provided the best number of calories is even less likely. This means that a change *away* from Stone-Age conditions might improve health.

Before I made the observations described in Example 1, I rarely thought about mismatch ideas. When I did, I considered them inevitably post hoc – used to explain results already obtained – and too vague to be helpful. Example 1 made me reconsider. Self-experimentation was so easy that one could test strange ideas and ideas likely to be wrong. Making my life slightly more Stone-Age was often possible.

2.3.2. Birth of the idea. What mismatch might cause early awakening? Based on observations of people living in temporal isolation, Wever (1979) concluded that contact with other people has a powerful effect on when we sleep and theorized that a circadian oscillator responds to human contact so as to make us awake at the same time as those around us. I believe that events that affect the phase of a rhythm usually affect its amplitude, so Wever's conclusion, if true, suggested that human contact probably influences the amplitude of an oscillator that controls sleep. Early awakening could be due to too-shallow sleep. Anthropological field work suggested that Stone-Age people had plenty of social contact in the morning – for example, gossiping and joking and discussing plans for the day (e.g., Chagnon 1983, pp. 116–17). I lived alone and often worked alone all morning. Maybe the sleep oscillator needs a "push" each morning to have enough amplitude. If so, maybe lack of human contact in the morning caused early awakening.

My interest in a weight/sleep connection (Example 1) had led me, via Webb (1985), to a survey of thousands of adults in 12 countries about the timing of daily activities (Szalai 1972). Two of its 15 data sets (three countries had two data sets) were from urban areas in the United States; the rest were from urban areas in other industrialized countries (e.g., Belgium, Germany, Poland). Looking at the results, I noticed that Americans were much more often awake around midnight than persons in any other country (upper left panel of Fig. 3). The data in Figure 3 are from employed men on weekdays, but other subsets of the data showed the same thing. Only one other activity so clearly distinguished the United States from the other countries: watching television at midnight (upper right panel). Americans watched late-night television far more than did people in any other country. Americans did not eat unusually late (lower left panel). Americans worked later than people in other countries (lower right panel) but the difference seemed too small to explain the difference in wakefulness. The researchers concluded "it is not unlikely that television has something to do with [the] later retiring . . . times [of Americans]" (Szalai 1972, p. 123). TV watching resembles human contact in many ways. Perhaps late-night TV viewing affected Wever's person-sensitive oscillator.

If watching television can substitute for human contact in the control of sleep, it was easy to test the idea that my residual early awakening was due to lack of human contact in the morning. One morning in 1995, I got up at 4:50 a.m. and a few minutes later turned on the TV. During the period 1964–1966, which is when the multicountry time-use survey was conducted, the most popular TV program in

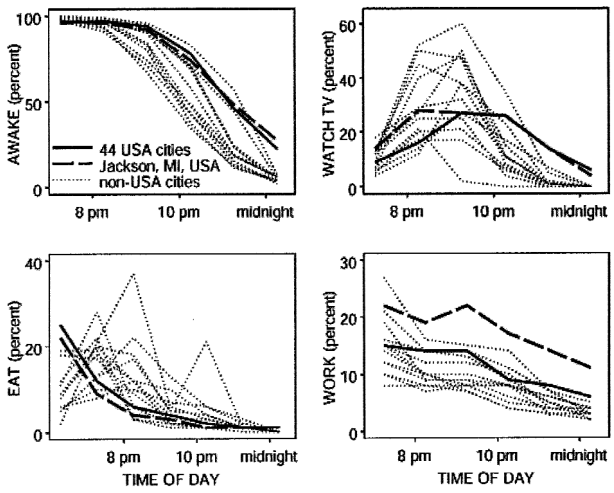

Figure 3. Probability of engaging in various activities as a function of time of day. The data are from employed men on weekdays. The thick lines show the two United States samples; the thin lines show the 13 samples from the other 11 countries. From Szalai (1972, pp. 746, 764, 766, 769).

America at 11 p.m. was *The Tonight Show,* which began with Johnny Carson's monologue. So that morning in 1995 I watched (on tape) Jay Leno's and David Letterman's monologues, which together lasted about 20 minutes. They were mildly amusing. I fell back asleep at 6:10 a.m. and awoke for good at 8:12 a.m. The day was uneventful. I went to bed at an ordinary time (11:49 p.m.).

I got up at 5:01 a.m. the next morning. To my astonishment, I felt great – cheerful and calm, yet full of energy. I could not remember ever feeling so good early in the morning. The good mood continued throughout the day.

What caused it? Nothing wonderful had happened the day before. Maybe I had slept well, but that would not explain why I felt so cheerful. Rare events are often caused by rare events, so I searched my memory for unusual events the previous day. All I could come up with was that I had watched television early in the morning. I had never before done that. If this was indeed the cause of my good mood, it meant that the Leno and Letterman monologues had improved my mood *the next morning,* one day later, in spite of having had no noticeable effect the same day.

2.3.3. Tests This conclusion was hard to believe, but easy to test. Over the next several days I confirmed it: Morning TV viewing made me feel better the next day but not the same day. An experiment with an ABA design (A = TV, B = no TV) showed clearly the mood change and the day-long lag between cause and effect (Roberts & Neuringer 1998).

Because of its big effect on mood, I assumed that the proper "dose" of morning TV viewing would eliminate early awakening. But months of trying different viewing possibilities (different starting times, different durations, different programs) did not produce improvement. Finally I gave up and turned to studying the mood effect itself.

Necessary and sufficient stimuli. At first, I believed that any TV watching would work. So I watched shows I enjoyed, such as *The Simpsons, Beverly Hills 90210, Melrose Place, Newsradio,* and Jay Leno's *Tonight Show* monologue. The results were far from uniform. My mood, although always positive, varied considerably from day to day, without obvious explanation.

To try to understand and reduce the variation, in July 1995 I started to measure and record my mood each day. I tried using the Profile of Mood States, a standard measure of mood, but even the short version (Shacham 1983) was too time-consuming. So I made my own scales, designed to be quick and sensitive to the obvious effects of morning TV viewing – namely, that I felt happier, less irritable, and more eager to do things. To record my mood, I wrote down three numbers. The first was a rating on the dimension unhappy/happy, with 5 = extremely unhappy, 10 = very unhappy, 20 = quite unhappy, 25 = unhappy, 30 = somewhat unhappy, 40 = barely unhappy, 50 = neither unhappy nor happy, 60 = barely happy, 70 = somewhat happy, 75 = happy, 80 = quite happy, 90 = very happy, and 95 = extremely happy. Ratings between these numbers (e.g., 46, 79) were permitted. The second number measured the dimension irritable/serene, with the same modifiers: 5 = extremely irritable, 10 = very irritable, 20 = quite irritable, and so on, up to 90 = very serene, and 95 = extremely serene. The third number measured the dimension reluctant/eager, with 5 = extremely reluctant, 10 = very reluctant, and so on, up to 95 = extremely eager. The three numbers were very highly correlated.

I continued to watch my favorite television programs for an hour or more – almost always several different programs. Viewing documentaries or *The Real World* (with the camera style of a documentary) was less effective in increasing my mood than other programs; stand-up comedy was more effective. Was humor the crucial ingredient? Probably not, because *The Simpsons* (very funny) seemed to have no effect. Another feature distinguishing stand-up comedy and documentaries was the proportion of time the viewer sees a large face looking at the camera: high for stand-up comedy, low for documentaries. This would explain why *The Simpsons* was ineffective: no large faces looking at the camera. Stand-up comedy also showed more faces and was more effective in producing a better mood than were the sitcoms and dramas I usually watched; documentaries showed fewer faces and were less effective than what I usually watched. These observations suggested that faces were necessary. The O. J. Simpson trial, full of faces in profile, had little effect, so apparently the faces had to be looking at the camera. Playing racquetball in the early morning, an activity in which I saw my opponent's face only a small fraction of the time, had little effect. This strengthened my conclusion that full-face views were necessary.

Were faces sufficient? If so, the duration of the nonface portion of any program should have no effect. I used results from December 1995 through June 1996, the first period for which the analysis could be done, to see if this was true. During this period, several experiments showed that face duration *was* important: 0 and 20 minutes of faces produced a lower mood the next day than did 30 minutes or 40 minutes of faces, which had roughly the same effect (Roberts & Neuringer 1998). It was possible to assess the effect of nonface TV viewing because what I watched varied considerably in face density (face duration/total duration). The programs I watched fell into two categories: those I enjoyed – such as *Friends* (face density 20–40%), *ER* (30–40%), and *Murder One* (40–60%) – and those with a high face density, mostly *Charles Grodin* (70–90%), *Charlie Rose* (60–80%), and *Rivera Live* (50–70%). (See the Appendix for more details.) Each morning I usually watched programs in both categories, but the proportion of

each varied a lot from day to day. I watched until I accumulated whatever face duration I was aiming for (usually 30 min). Thus, although the total time I spent watching television varied substantially, the duration of faces I saw was relatively constant.

More nonface TV viewing did not raise next-day mood – if anything, it lowered it (Fig. 4). The changes in nonface duration were not confounded with changes in face duration. When nonface duration was less than 40 minutes, median face duration was 30.2 minutes; when nonface duration was 40 minutes or more, median face duration was 30.1 minutes. The slight mood decrease at longer durations suggested that greater face density (watching *less* television but keeping total face duration constant) might raise mood. Indeed, an experiment that deliberately varied face density found this result (Roberts & Neuringer 1998).

Time course. I slowly realized that morning TV was not only raising my mood the next day (24 hrs later) but also lowering my mood that evening (12 hrs later). In 1999, I did an experiment with an ABABA design to map the time course of the effect. The A (faces) and B (no faces) conditions were the same except that during B the upper two-thirds of the TV screen was covered.

The TV had a 27-inch (68-cm) screen (measured diagonally). During all phases of the experiment I started watching at about 6:00 a.m. (median 6:04 a.m., range 6:01–6:15 a.m.) and continued watching until I had watched a total of 60 minutes of "big faces" – faces that were roughly life-size with both eyes visible. To qualify, the shoulder width of the person being viewed had to fill 60% to 80% of the screen width. I kept track of the accumulated face duration using a stopwatch. Viewing stopped at a median time of 7:31 a.m. (range 7:17–7:45 a.m.). About half of the time was spent watching *Booknotes* and *Washington Journal;* the other half was spent watching *Charlie Rose, The Larry King Show,*

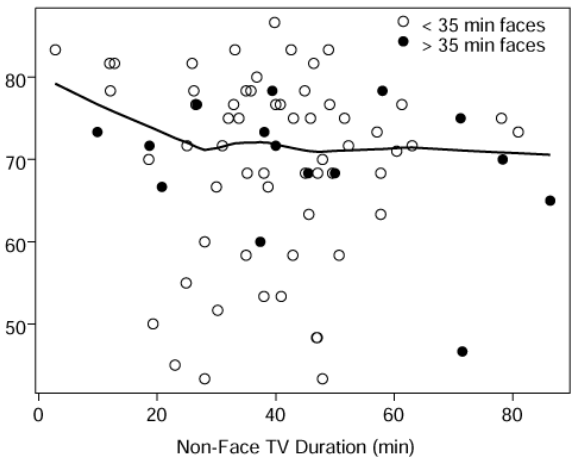

Figure 4. Next-day mood as a function of duration of viewing nonface television. Nonface TV duration = Total TV duration − Face duration. Face durations ranged from 25 to 45 minutes; unfilled points indicate that face duration was less than 35 minutes, filled points indicate more than 35 minutes. The TV viewing started between 6 a.m. and 8 a.m. Mood ratings were made from 10 a.m. to 2 p.m. the day after viewing. Each point is an average of three ratings, one for each scale. The three scales measure the dimensions unhappy/happy, irritable/serene, and reluctant/eager (scale range: 5 = extremely negative [e.g., extremely unhappy], 95 = extremely positive [e.g., extremely happy], with 50 = neither negative nor positive [e.g., neither sad nor happy]).

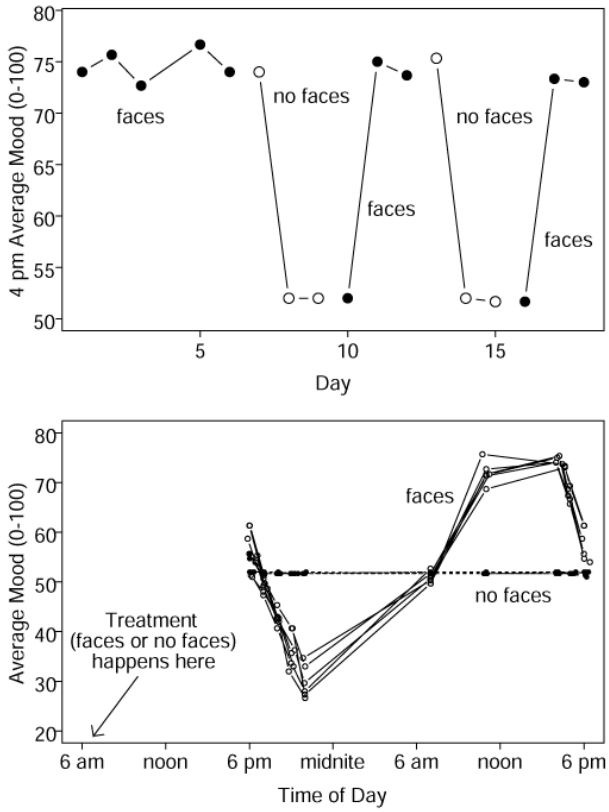

Figure 5. Mood ratings over 17 days, beginning October 13, 1999. Upper panel: Mood at 4:00 p.m. day by day. Lower panel: Time course of the effect. In both panels each point is an average of three ratings, one for each scale. The three scales measure the dimensions unhappy/happy, irritable/serene, and reluctant/eager (scale range: 5 = extremely negative, 95 = extremely positive, with 50 = neither negative nor positive). Each line is a separate series of measurements. The data start about 12 hours after the treatment because that is when the treatment began to have an effect.

The O'Reilly Factor, The Today Show, and the talking-heads segments of *The Newshour with Jim Lehrer.* With the exception of *The Today Show* and some of *Washington Journal,* all of this was taped, and I usually skipped portions without useful faces. My eyes were about 1 meter from the screen. In the evenings after watching television in the morning, I avoided seeing faces after 7:00 p.m. (One evening this was unavoidable, and data from the next day were excluded.)

I recorded my mood on the rating scales at about 7:00 a.m., 11:00 a.m., and 4:00 p.m. for Days 2–18; I also recorded it every hour after 4:00 p.m. until I fell asleep. The upper panel of Figure 5 shows mood at 4:00 p.m. As usual, seeing faces raised my mood with a one-day lag. The lower panel of Figure 5 shows that exposure to faces produced an oscillation in mood that lasted about 24 hours, starting about 12 hours after I saw the faces.

Figure 6 emphasizes when the face-viewing effect started and stopped. Both panels show when it started. Both show that whether or not I watched faces in the morning made no difference until about 6 p.m. that evening. The lower panel shows when the effect stopped. When I had seen faces the previous morning (Sunday morning), the lower panel shows that their effect stopped at about 6 p.m. the next day (Monday evening) in the sense that the oscil-

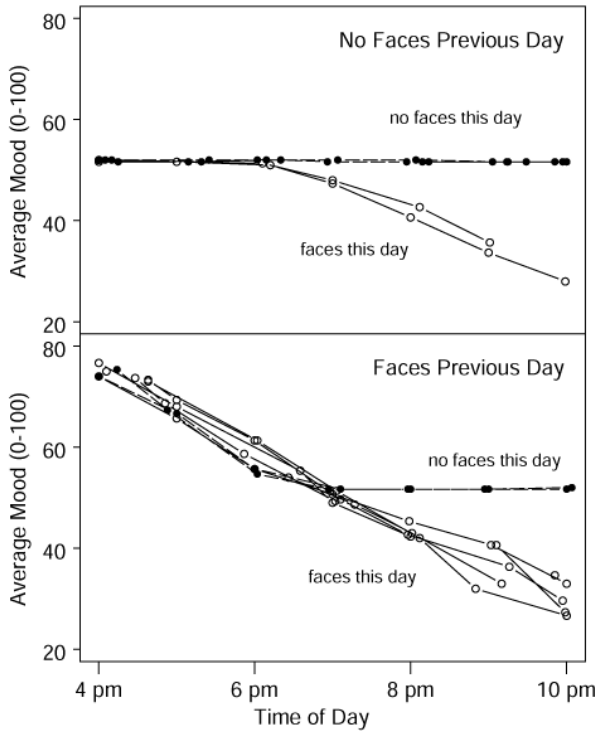

Figure 6. When the face effect started (upper panel) and stopped (lower panel). Both panels: Mood as a function of time of day and whether faces were viewed that morning. Upper panel: From days when no faces were viewed the previous day. Lower panel: From days when faces were viewed the previous day. Each point is an average of three ratings, one for each scale. The three scales measure the dimensions unhappy/happy, irritable/serene, and reluctant/eager (scale range: 5 = extremely negative, 95 = extremely positive, with 50 = neither negative nor positive).

lation ceased – the line became flat. Another way to see when the effect ends is to compare the upper and lower panels. They differ before 6 p.m. but not after.

Visual features. Distance mattered. Faces on television had substantially less effect when I viewed them 2 m from the screen than when I was 1 m away (Roberts & Neuringer 1998).

Comparison of 20-, 27-, and 32-in. (51-, 69-, and 82-cm) TV screens (Fig. 7) showed that face size matters and that the 27-in. screen was best. I began with 8 days of viewing a 20-in. TV screen from a distance of 1 m (for 30 min). On the following 5 days I viewed a 27-in. screen, also from a distance of 1 m and for 30 minutes. My mood increased, $t(11) = 4.16$, $p = 0.002$, after the change to the larger screen. Then I increased face time to 50 min/day and made more comparisons. An ABCDA series of treatments (Days 26–104), with A = 27-in. screen viewed from a distance of 1 m and B = 20-in. screen viewed from a distance of 0.75 m (instead of 1 m, to roughly equate retinal size), C = 32-in. screen at 1 m, and D = 32-in. screen at 1.5 m, found that mood was lower with the 20-in. than with the 27-in. screen (B vs. both A's combined), $t(27) = 2.35$, $p = 0.03$, and lower with the 32-in. screen than with the 27-in. screen (C vs. both A's combined), $t(31) = 8.53$, $p < 0.001$. Viewing the 32-in. screen from farther away reduced mood ($t(15) = 8.29$, $p < 0.001$) rather than increasing it. So it was not retinal size that mattered.

An ABCA series (Days 77–170) done to check these conclusions found the same results. Viewing a 20-in. screen (B) produced a lower mood than did viewing a 27-in. screen (A), $t(35) = 7.69$, $p < 0.001$, and viewing a 32-in. screen (C) produced a lower mood than did viewing a 27-in. screen (A), $t(40) = 10.64$, $p < 0.001$.

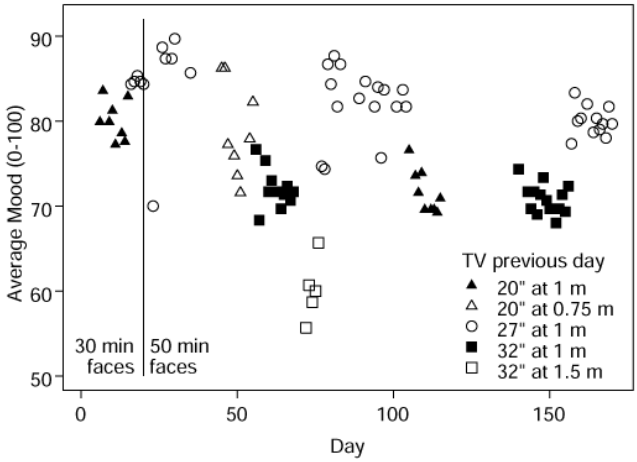

Figure 7. Effect on mood of TV size and distance. The vertical line separates results the day after viewing 30 minutes of faces from results the day after viewing 50 minutes of faces. Day 1 = November 17, 1996. Average mood = mood at 4 p.m. (average of three ratings on three scales measuring different dimensions of mood).

The results also suggested (again) that distance mattered. With a 20-in. TV screen, mood was lower when the screen was viewed from a distance of 1 m than when it was viewed from 0.75 m, $t(15) = 3.55$, two-tailed $p < 0.006$ – although the decline over days in the baseline condition (27-in. screen at 1 m) makes this difference hard to interpret. With a 32-in. screen, mood was higher when the screen was viewed from a distance of 1 m than when it was viewed from 1.5 m, $t(15) = 8.29$, $p < 0.001$, a change opposite to the downward baseline trend.

To give an idea of the size of the faces I watched, the face length from the bottom of the chin to the top of the head (not counting hair) was usually about 50% of the vertical length of the screen. A 27-in. screen made these faces closer to life-size than did 20-in. and 32-in. screens. The distance and face-size results showed that the most potent stimulus resembles what is seen during a conversation: a life-sized face about 1 m away.

Time of day. A one-hour change in time of day of viewing made an easily detectable difference (Roberts & Neuringer 1998). The upper panel of Figure 8 shows a repetition of this result. Viewing television beginning at 6 a.m. produced a lower mood the next day than viewing beginning at 7 a.m., $t(27) = 11.08$, $p < 0.001$.

In 1996, I noticed several times that social contact during the evening (e.g., from a dinner) was followed by a lower-than-usual mood the next day. The first few instances were surprising but they were followed by dozens of confirmatory instances. Eventually I did an experiment. The lower panel of Figure 8 shows the design and results. Between baseline phases (50 min of faces starting at 4:30 a.m., no evening TV viewing), I added 10 minutes of faces starting at various evening times (10:00 p.m., 8:00 p.m., 6:00 p.m.) to the baseline level of viewing. Watching television at 10 p.m. reduced mood, $t(18) = 15.64$, $p < 0.001$. In addition, watching at 8 p.m. reduced mood, $t(18) = 9.82$, $p <$

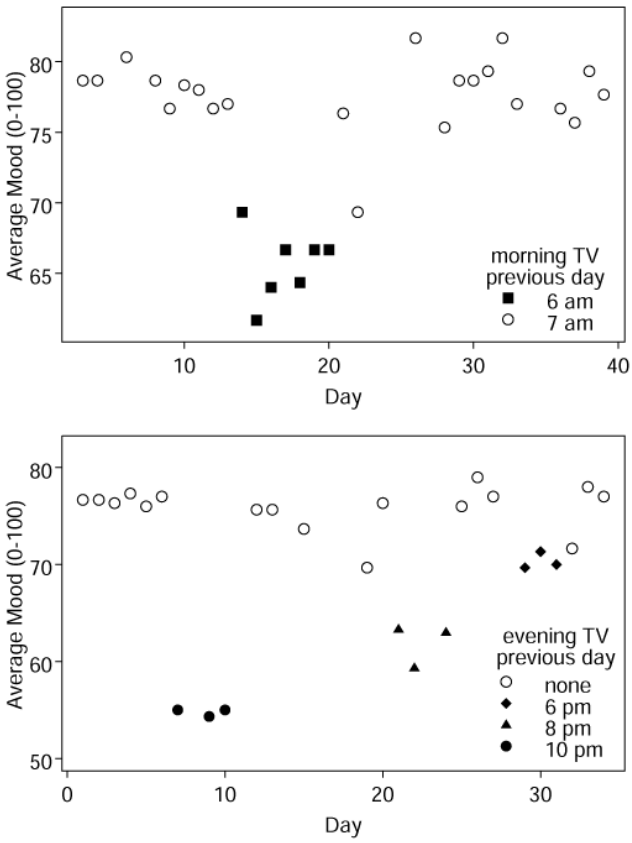

Figure 8. Effect on mood of time of day that faces are seen. Upper panel: Effect of the time that faces are watched on television in the morning; 6 a.m. and 7 a.m. are the starting times. Day 1 = May 11, 1997. Lower panel: Effect of watching faces on television in the evening; 6 p.m., 8 p.m., and 10 p.m. are the starting times. Day 1 = October 6, 2000. Average mood = mood at 4 p.m., average of three ratings on three scales measuring different dimensions of mood.

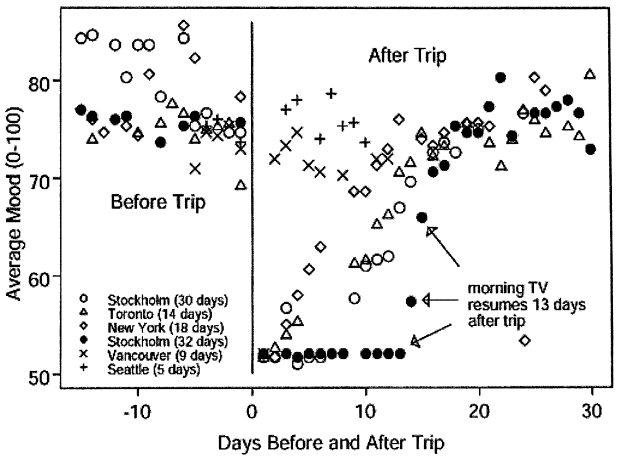

Figure 9. Effect on mood of long trips. The trips are listed in the order in which they occurred. Data include up to 15 days before the trip and 30 days afterwards, but only those days for which the treatment was the same. For example, suppose a trip started on Day 100 and ended on Day 110, and the treatment (the details of TV watching, such as the starting time and duration) remained the same over the periods Days 90–99 and Days 111–120. Then Days 91–99 and Days 111–121 would be shown. Data were not included from days *after* (a) no TV viewing or later viewing than usual or (b) I had face-to-face contact after 7 p.m. (e.g., if I had such contact on Monday, Tuesday's data were not included). The filled points indicate a trip after which I waited 13 days before I resumed morning TV viewing. Average mood = mood at 4 p.m., average of three ratings on three scales measuring different dimensions of mood.

0.001. To my surprise, watching at 6 p.m. also reduced mood, $t(18) = 4.07$, $p < 0.001$. The four baseline phases did not differ, $F(3, 13) = 1.81$, $p = 0.20$.

Time-of-day effects (Fig. 8) suggest the existence of an internal circadian clock that varies sensitivity to faces over the course of the day. Such a clock must be "set" by the outside world, with sunlight the likely candidate to perform that function. This seemingly obvious conclusion eluded me, however, and I was quite surprised when I returned home from Europe one day in 1997 (my first long trip after discovery of the effect) to find that the effect had vanished: Morning TV viewing did not raise my mood. During the months before the trip, the effect had been completely reliable, reappearing immediately after a day without morning TV watching (e.g., upper panel of Fig. 5). Over the next three weeks, the effect gradually returned. In contrast, irregular sleep and sleep difficulties went away after a few days. Figure 9 shows the mood ratings before and after that trip (a 30-day trip to Stockholm) and several later trips. All trips to the East Coast and Europe made the effect disappear for about three weeks, whereas North-South trips (to Vancouver and Seattle) had no detectable effect. After a later trip to Europe, I waited 13 days before resuming the morning TV watching. The effect reappeared at near-full strength almost immediately. Because a long trip causes many changes, there are many possible explanations of the initial observation that a long trip caused the effect to disappear. But most of them were ruled out by the later results (no effect of North-South travel, quick reappearance after delayed resumption). All of the results are consistent with the explanation that the changes in the timing of light produced by east-west trips disrupted the oscillations of a light-sensitive circadian oscillator.

2.3.4. Related results. In several ways, the results make sense in terms of prior knowledge.

(a) *Symptoms of depression.* The obvious effects of morning faces on my mood – I felt happier, more serene, and more eager to do anything – involved the same dimensions that change (in the opposite direction) during depression. A *DSM-IV* (*Diagnostic and Statistical Manual of Mental Disorders* 1994) diagnosis of major depression requires lack of happiness ("depressed mood," p. 327) and lack of eagerness ("markedly diminished interest or pleasure in all, or almost all, activities," p. 327). Irritability is common. For example,

> [Reluctance:] Everything was too much of an effort. She sat and stared, and did not want to be bothered with anything. . . . [Irritability:] She became very harsh with the children and shouted at them and hit them (most unusual for her). (Brown & Harris 1978, p. 308)

See "Causes of depression" in section 2.3.5 for more about the similarity.

(b) *Bipolar disorder.* Bipolar disorder is distinguished by the presence of both mania and depression (*DSM-IV* 1994). Mania may include irritability, but in other ways corresponds to being high on the dimensions unhappy/happy and reluctant/eager (*DSM-IV* 1994). In some cases, the pa-

tient goes back and forth between mania and depression at roughly equal intervals, which has often been taken to imply that an oscillator controls mood (e.g., Checkley 1989; Georgi 1947; Halberg 1968; Kripke 1984). Pettigrew and Miller (1998) found that persons with bipolar disorder had a slower rate of switching between percepts in a binocular rivalry task than did normal controls, which likewise suggests that an oscillator controls mood.

(c) *Connections between depression and circadian rhythms* (Checkley 1989; Hallonquist et al. 1986; Wehr & Goodwin 1983). Depression is strongly correlated with circadian abnormalities, especially insomnia (Bixler et al. 2002; Breslau et al. 1996; Chang et al. 1997; Ford & Kamerow 1989; Ohayon 2002). Compared with other psychiatric patients, patients with affective disorders (mostly depression) had the most disturbed sleep (Benca et al. 1992). In addition, circadian-rhythm manipulations, such as sleep deprivation, sometimes alleviate depression (Bunney & Bunney 2000). Finally, depression often shows a large circadian rhythm: worse in the morning, better as the day wears on (Haug & Fähndrich 1990; Moffoot et al. 1994).

(d) *Connections between mood and circadian rhythms.* Many studies have observed a circadian rhythm in mood in normal subjects (e.g., Boivin et al. 1997; Stone 1996). Totterdell et al. (1994) found that, within subjects, earlier sleep onset predicted better mood the next day and "was a better predictor [of mood] than sleep duration" (p. 466). The better mood included higher ratings of cheerfulness, which corresponds to what I called unhappiness/happiness, and better ratings of "social interaction experience," which may correspond to irritability/serenity. Totterdell et al. (1994) reviewed other findings that suggest connections between a circadian mechanism and mood.

(e) *Social entrainment of circadian rhythms.* The tendency for humans or animals living together to become synchronized has been well-established in animals (Rajaratnam & Redman 1999). The most persuasive human evidence comes from Wever (1979), who studied 12 groups of two subjects living in temporal isolation. Vernikos-Danellis and Winget (1979) and Apfelbaum et al. (1969) found similar results studying slightly larger groups.

(f) *Correlation between isolation and depression.* Persons who live alone or who are socially isolated are more likely to have symptoms of depression (Kaplan et al. 1987; Kennedy et al. 1989; O'Hara et al. 1985; Swenson et al. 2000).

2.3.5. Discussion. The effects on mood of viewing faces cannot be due to expectations, of course. The first good mood was a surprise. A 12-hour gap between cause and effect (Fig. 6) is unprecedented in the study of mood. Even after the effect of morning faces became expected, surprises continued (the effect of travel, Fig. 9; the effect of 6 p.m. TV viewing in Fig. 8).

2.3.5.1. Mechanistic explanation. The broad outlines of an underlying mechanism are shown in Figure 10. A light-sensitive circadian oscillator (Fig. 9) determines "sensitive" times of day (Fig. 8) during which the sight of faces (Fig. 4) will cause a down-up oscillation in mood (Fig. 5) that begins about 12 hours later and lasts about 24 hours (Fig. 6). Faces viewed at night also affect the face-sensitive oscillator, tending to cancel the effect of morning viewing (Fig. 8). That faces viewed at night should affect the oscillator is only

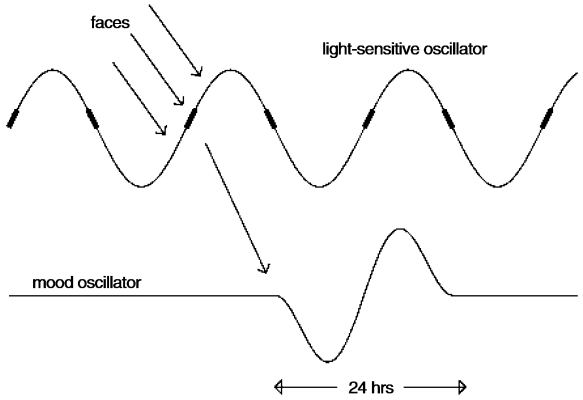

Figure 10. Explanation of the effect on mood of viewing faces in the morning. Faces act through a gate, which opens in the morning and evening, controlled by a light-sensitive circadian oscillator. Signals that pass through the gate produce one cycle of a circadian oscillation in mood.

to be expected, given that its function is to synchronize sleep and moods (discussed below). I measured the effect of night faces only on amplitude; but presumably they changed the phase of the mood oscillation as well.

Travel across time zones reduces the effect because it flattens (reduces the amplitude of) the light-sensitive rhythm. As the light-sensitive oscillator adjusts to the new time zone, the amplitude recovers. This creates a problem for replication attempts because it means that morning and evening exposures to sunlight need to have been roughly constant for a few weeks to observe a large effect of face viewing. Suppose a person is usually indoors until 9 a.m. If he happens to leave home at 7 a.m. and is outside for several hours, the light-sensitive oscillator may change as much as if the person had traveled across two time zones. Fluorescent light at ordinary indoor intensities can affect a light-sensitive circadian oscillator (e.g., Boivin et al. 1996; Waterhouse et al. 1998; Zeitzer et al. 2000, studies which all used cool white fluorescent lamps), so this too must be controlled. Incandescent light, however, is unlikely to cause problems. It is much redder than cool-white fluorescent light, and the circadian oscillator that controls activity is much more sensitive to blue-green light than to reddish light (Morita & Tokura 1998; Takahashi et al. 1984).

2.3.5.2. Evolutionary explanation. The effects studied here imply a mechanism that tends to produce synchronized mood cycles in persons living together, assuming that they chat in the morning. Sight of a face (both eyes) is a shared event: When I see your face, you see my face. Other forms of contact (voice, sight of other body parts) are less reciprocal. I can hear your voice without you hearing mine, for example. Touch is reciprocal but not at all restricted to social contact. This suggests why seeing a face is the crucial event.

Synchronized mood cycles were beneficial, it can be argued, because they solved problems associated with group living and group work.

A rhythm of *serenity/irritability* protects sleep by keeping others quiet while you are sleeping. Group living makes it inevitable that some will be asleep while nearby persons are awake. Irritability at night, while sleeping, means that if someone wakes you up you will snap at him, making him

less likely to do it again. Such experiences teach awake persons to be quiet while others are sleeping. A *rhythm* of irritability (high at night, low during the day) is necessary because high irritability during the day would make cooperation more difficult.

A rhythm of *eagerness/reluctance* synchronizes sleep. The more reluctant you are, the less active you are. To fall asleep requires inactivity; you stop moving, lie down, and soon fall asleep. So making everyone inactive at the same time tends to synchronize sleep as well. Synchronized sleep promotes working together (because cooperation usually requires that co-workers be awake at the same time) and reduces noise during sleep. Less noise makes it easier to sleep, of course, and allows sleepers to adopt a lower threshold for waking up due to sound, sound that would probably be caused by an intruder.

Finally, a rhythm of *happiness/unhappiness* also promotes working together. This argument rests on two assumptions: that happiness reduces complaints and that complaints make group work more difficult. Happy people complain less than unhappy people; the correlation between happiness and absence of complaints is so clear that the meanings overlap (*Are you happy?* often means *Are you complaint-free?*). That happy people complain less than unhappy people can be explained if we assume that unpleasant internal states (heat, thirst, hunger, boredom) reduce the level of happiness, but only when the level goes below neutral (you become unhappy, not just less happy) will you try hard to fix the problem (perhaps by complaining). Happy people have a buffer – a "reserve" of happiness – that unhappy people do not. Suppose, for example, that a certain level of thirst reduces your level of happiness by 10 points on a scale where 50 is neutral. If you start at 70, this amount of thirst moves you to 60. Still above 50, you do not complain. But if you start at 50, this level of thirst moves you to 40, and you complain and/or stop working until your thirst is satisfied. The assumption that complaints make group work more difficult derives from the belief that if several people are working together and one of them complains about something, work may have to stop while the complaint is dealt with. Happiness in the late morning and afternoon facilitates group work during those times by reducing complaints during them. The rest of the day remains for solving the problems (indicated by hunger, thirst, and so on) that midday happiness kept hidden.

2.3.5.3. Causes of depression. The self-experimentation work described here did not directly involve depression (i.e., I was not depressed), but for several reasons seems likely to help us understand it.

First, there is the similarity of the dimensions affected (mentioned earlier). Viewing faces made me unhappy, irritable, and reluctant and then later happy, serene, and eager (lower panel of Fig. 5). The core symptoms of depression are unhappiness and reluctance, and irritability is common. The similarity of dimensions is impressive because (a) I was not depressed and (b) the combination of changes is counterintuitive. Simultaneous unhappiness and reluctance is counterintuitive because unhappiness is close to dissatisfaction, which implies greater chance of action than usual (to improve things), whereas reluctance implies less chance of action. Similarly, simultaneous irritability and reluctance is counterintuitive because irritability implies greater chance of action (in response to unpleasant things). No theory of depression, as far as I know, explains why these symptoms are found together (Schwartz & Schwartz 1993).

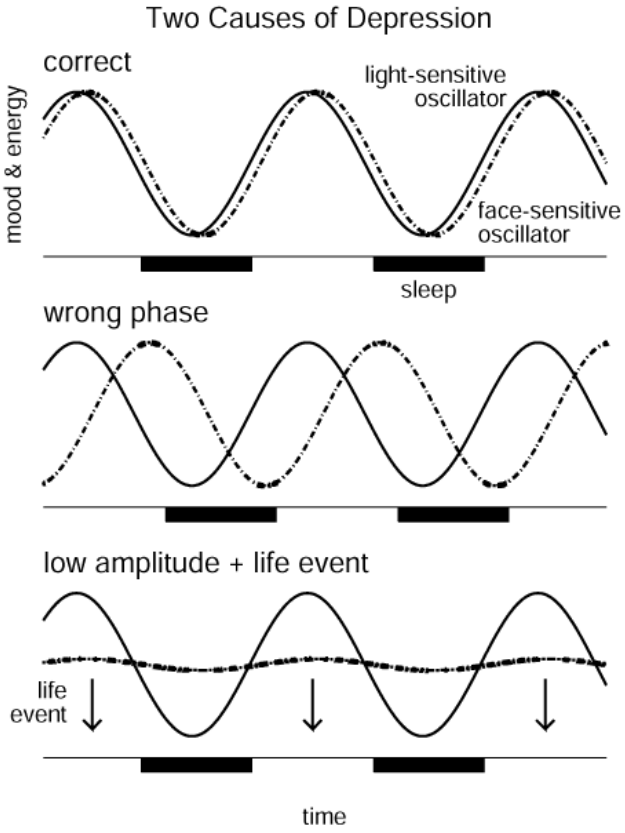

Figure 11. Two ways that a malfunctioning mood oscillator could cause depression: wrong phase or low amplitude. *Energy* means energy in the everyday sense used to describe human behavior (e.g., "full of energy"); a technical definition would be wakefulness.

Second, a simple theory based on the results readily explains some basic facts about depression. Let us assume that the timing of sleep is controlled by two oscillators: the light-sensitive oscillator revealed by thousands of circadian-rhythm experiments and the face-sensitive oscillator revealed here (Fig. 11). Wakefulness (called *energy* in Fig. 11) depends on something like the sum of the two rhythms. The system is designed to work well when both have high amplitude and are nearly in phase, as illustrated in the top panel of Figure 11.

The face-sensitive oscillator, which controls mood as well as sleep, can, like any oscillator, malfunction in two ways: wrong phase or low amplitude. *Wrong phase* (illustrated in the middle panel of Fig. 11) may be caused by exposure to faces at night (via TV viewing or via socializing). Faces at night would shift the peak of the mood rhythm toward later times, delaying sleep. But because the face oscillator has only partial control of sleep, sleep times would shift less than the mood rhythm. For instance, a 6-hour mood-rhythm shift might delay sleep times only 2 hours. The incomplete displacement of sleep times causes the low part of the mood cycle to be "uncovered" so that mood is low upon awakening. Mood improves while awake. The two predictions of this pattern are (1) later-than-usual sleep times, and (2) mood should be lowest in the morning and improve throughout the day.

Indeed, depression has been strongly associated with

staying up late (Regestein & Monk 1995) and with insomnia (see the references cited above). Morning-to-evening improvement is common (e.g., Haug & Fähndrich 1990; Moffoot et al. 1994) and hard to explain. "The various psychological, cognitive and biological hypotheses of depressive illness tend to avoid the issue," noted Moffoot et al. (1994, p. 258).

Even more puzzling has been that depression, unlike any other common disorder, can be dramatically reduced by treatments that would seem likely to cause nothing but harm – namely, electroconvulsive therapy (ECT) and sleep deprivation. The phase-shift theory of Figure 11 offers an explanation for the effectiveness of these treatments because, as I discovered after east-west trips (Fig. 9), the face/mood effect requires something fragile: the light-sensitive rhythm that gates the effect of faces (Fig. 10). If this rhythm is flattened, the effect of faces – positive (as in my research) or negative (as, I hypothesize, in many cases of depression) – is eliminated. Obviously ECT could flatten this rhythm. Sleep deprivation could flatten this rhythm because the subject is exposed to fluorescent light (in the lab) all night. Travel across time zones should be equally effective. Solomon (2001), whose depression showed morning-evening improvement (pp. 54, 63), traveled to Turkey (from New York, apparently) and found that "in the perfect Turkish sunshine the depression evaporated" (p. 78).

Low amplitude of the oscillator (illustrated in the bottom panel of Fig. 11) may be caused by fluorescent light at night (if irregular and strong enough), by too little exposure to morning faces, or by too much exposure to evening faces (canceling the effect of morning faces). A low-amplitude mood oscillation does not by itself cause depression; but it leaves a person vulnerable to mood-lowering events, such as divorce or death of a spouse. The event that pushes you from 80 (quite happy) to 50 (neither happy nor sad) when your mood oscillator is working properly will push you from 50 (neutral) to 20 (quite unhappy) when it is not. Indeed, a large fraction of cases of depression seem to be triggered by a serious mood-lowering event, such as a major loss (e.g., Brown 1998; Kessler 1997; Stueve et al. 1998). If mood-lowering events cause unhappiness and reluctance but not irritability, this would explain why many depressed persons are not irritable.

The difference between a depression that is triggered by difficulties, and one that is not (see the bottom and middle panels, respectively, of Fig. 11), is the same as the difference between reactive and endogenous depression, and the evidence for such a distinction is "substantial" (Nelson & Charney 1981, p. 2). Research about this distinction has asked whether the presence or absence of diurnal variation (another difference between the two types of depression in Fig. 11) predicts anything important. Many studies have found that it does (Carpenter et al. 1986; Haug & Wirz-Justice 1993). For example, "sleep deprivation responders manifest a higher percentage of diurnal variation than non-responders, with a pattern of improved mood in the evening" (Haug 1992, p. 271).

Third, the theory of Figure 11 can easily explain, in addition, some of the most important epidemiology. Studies of depression have found three very large risk ratios (10 or more):

1. *Amish versus America.* Over a five-year period, only 41 cases of unipolar depression were identified among 12,500 Pennsylvania Amish, which is a one-year rate of about 0.1 percent (Egeland & Hostetler 1983). In contrast, a national survey found that the one-year rate of major depressive episodes in the United States was 10 percent (Kessler et al. 1994) – 100 times greater. Because the two surveys used different diagnostic criteria the rates are not exactly comparable, but the difference is so large that it is hard to avoid concluding that the Amish lifestyle has a very strong protective effect.

Explanation: The Amish do everything right. Amish workdays begin "between four and five o'clock in the morning . . . Breakfast [eaten communally] may be served at five-thirty during the busy season of the year . . . Bedtime comes at nine o'clock or a little later" (Hostetler 1980, p. 136). The communal breakfast means lots of conversational exposure to faces in the morning; the early bedtime means no exposure to faces late at night. Amish houses and other buildings do not have fluorescent lights, thus there is no exposure to fluorescent light at night (Kraybill 1989). The average American has much less exposure to faces in the morning, much more exposure to faces (including televised faces) after 9 p.m., and much more exposure to fluorescent light at night.

2. *Chronic insomnia versus no insomnia.* In a community survey, Ford and Kamerow (1989) found that persons who reported severe insomnia at two interviews a year apart had a 40-fold increase in their likelihood of developing major depression between the interviews, compared to persons who reported severe insomnia at neither interview (Ford & Kamerow).

Explanation: The theory of Figure 11 assumes that the mood oscillator influences sleep. If this oscillator malfunctions, both mood and sleep will be affected. Whether the mood oscillator has the wrong phase (Fig. 11, middle panel) or low amplitude (bottom panel), there will be less pressure to sleep than when the system is working properly (top panel). Not enough pressure to sleep can cause difficulty falling asleep, difficulty staying asleep, and waking up too early, three types of insomnia.

3. *Now versus 100 years ago.* Depression rates apparently increased about 10-fold during the 1900s (e.g., Klerman & Weissman 1989; Lewinsohn et al. 1993).

Explanation: Several changes during the 1900s made it harder to benefit from morning faces. Electric lighting became common, making it easier to stay up late; and movies, television, telephone, and radio became widespread, providing more reason to stay up late. As people stayed up later, they saw more faces at night and woke up later, leaving less time for face-to-face contact before leaving for work, and face-to-face contact at work went down as jobs became more mechanical and electrical. The introduction of fluorescent lighting allowed exposure to fluorescent light at night, of course.

No well-known theory of depression easily explains any of these three risk ratios, as far as I can tell, although proponents of those theories might disagree.

2.3.5.4. Value of self-experimentation. Ehlers et al. (1988) proposed that depression is sometimes due to disruption of social rhythms. Healy and colleagues made similar proposals (Healy & Williams 1988; 1989; Healy & Waterhouse 1990; 1995). But as far as I know no one has managed, using only conventional research, to go from these ideas to something more specific. To connect circadian rhythms and mood makes considerable sense given the strong correla-

tion between depression and insomnia. However, the specific cause-effect linkage studied here was at first quite implausible because of the many hours between cause (seeing faces) and effect (mood change). It is hard to see how conventional research could have discovered it.

Even after the initial discovery, research with other subjects would have been very difficult. To study the effect using other subjects requires consistent repetition of the effect. The self-experimentation that followed the initial observation showed that the effect depended on many factors (TV size, viewing distance, and time of day, exposure to evening faces, exposure to sunlight, and, probably, fluorescent light) that can easily vary a lot from person to person. Before those factors were discovered and controlled, it would have been hard to consistently repeat the effect.

2.4. Example 3: Standing → Better sleep

2.4.1. Background. In 1996, early awakening remained a problem for me. Example 1 had suggested that it might be due to differences between my life and Stone-Age life. Example 2 had shown that this conclusion could be helpful. But I had no idea how to use these insights.

2.4.2. Birth of the idea. Around this time, I heard two stories about walking and weight loss. A friend who often went to Italy said that while visiting large cities he lost weight, but while visiting small ones he gained weight, presumably because he walked much more in large cities than in small ones. Another friend said that she lost a lot of weight on a month-long hike. Walking is a common treatment for obesity (e.g., Gwinup 1975), but in these two stories the amount of walking (many hours/day) was much more than in research studies.

Would walking many hours every day cause me to lose weight? I did not have the time to find out. I noticed a confounding, though: If you walk more than usual, you will stand more than usual, taking *standing* to include any time all your weight is on your feet. If walking causes weight loss, which seemed likely, it might be due to energy use or to standing. Does standing many hours each day cause weight loss? Probably not, I thought, but decided to try it anyway, partly because of my new interest (due to Examples 1 and 2) in a Stone-Age way of life. Stone-Age humans surely spent much more time on their feet than I did at the time (about 2–4 hrs/day).

From August 27, 1996 onwards, I began to stand much more. I raised the screen and keyboard of my office computer to standing level. I stood while using the phone. I walked more, bicycled less. The first few days were exhausting but after that it wasn't hard. I used the stopwatch on my wristwatch to measure how long I stood each day. I included any time that all my weight was on my feet: standing still, walking, playing racquetball. My weight did not change. (Maybe I lost fat, because my legs gained muscle.) Within a week, however, I realized I was awakening early less often. The improvement outlasted the initial exhaustion. During the three previous months, I had awakened too early on about 60% of days; during the first few months of standing, the frequency was about 20%.

2.4.3. Tests. How long I stood varied by several hours from day to day, mostly because of variations in events that required sitting (meetings, meals, travel). At first, I believed that any substantial amount of standing (e.g., 6 hrs) would reduce early awakening. In October 1996, however, I analyzed the data in preparation for a talk. The results of this analysis, shown in the lower panel of Figure 12, surprised me. Standing 5.0 to 8.0 hours had little effect on early awakening, judging from the pretreatment baseline (before August 27); standing 8.0 to 8.8 hours apparently reduced early awakening; and standing for 8.8 hours or more eliminated it – an unusual dose-response function. After that, I tried to stand at least 9 hours every day. This solved the problem. If I managed to stand at least 8.8 hours, I almost never awoke too early.

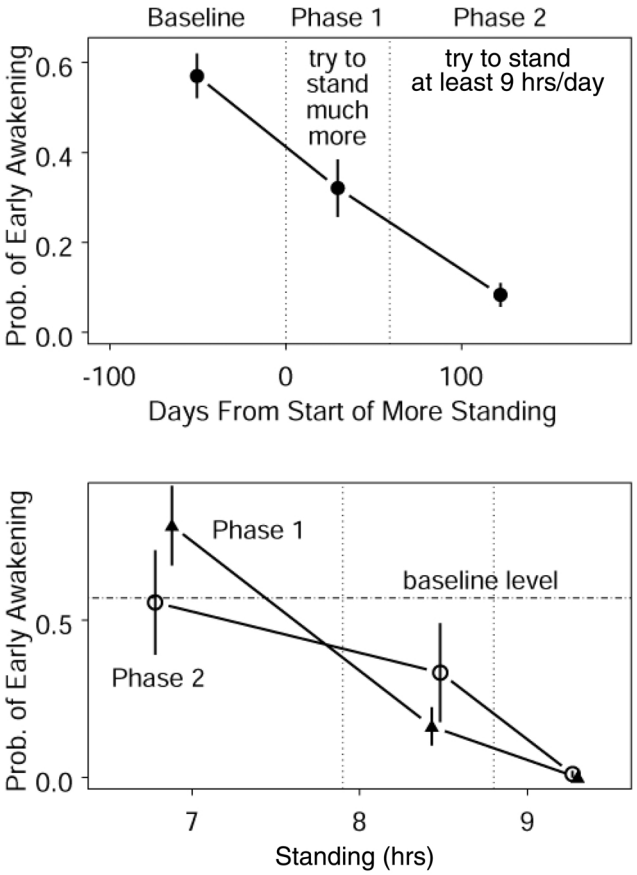

Figure 12. Effect of standing on early awakening. Early awakening = Fell back asleep within 6 hours after getting up. Vertical segments show standard errors computed assuming a binomial distribution. Because of travel and illness, some days were not included. Upper panel: Between-phase differences. Baseline = May 18, 1996–August 26, 1996. Phase 1 = August 27, 1996–October 24, 1996. Phase 2 = October 25, 1996–February 28, 1997. Lower panel: Within-phase differences. Standing durations were divided into three categories: fewer than 8.0 h; 8.0–8.8 h; and more than 8.8 hrs. The probability of early awakening for each category is plotted above the median of the durations in that category.

The upper panel of Figure 12 shows that the probability of early awakening decreased as I intentionally stood more. The overall likelihood of early awakening decreased from baseline to Phase 1, Fisher's test, two-tailed $p < 0.001$; and from Phase 1 to Phase 2, Fisher's test, two-tailed $p < 0.001$. The lower panel of Figure 12 shows the standing/sleep correlation within phases. As mentioned above, it was the discovery of this relation during Phase 1 that led to Phase 2. During Phase 1, the likelihood of early awakening de-

creased as standing increased from the shortest (5.0–8.0 hrs) to the longest (8.8–11.0 hrs) durations, Fisher's test, two-tailed $p = 0.04$. Phase 2 showed the same correlation, Fisher's test, two-tailed $p < 0.001$.

To make it easier to stand 9 hours per day, I bought two treadmills (office and home) so I could walk during computer use. (It is easier to walk for an hour than to stand still for an hour.) With their help, I often managed to stand 10 hours or more, and this led to another surprise: On several occasions, after standing that much, I awoke the next morning feeling completely rested. It was such a rare feeling that it reminded me of camping trips taken many years earlier. ("What does it feel like?" a friend asked. "Scrubbed clean of tiredness," I said.) It might seem obvious that a change that improves sleep in one way (less early awakening) may improve sleep in other ways (more rested when you awake). Nevertheless, it had not occurred to me. My experiences seemed to imply that the effect of 10 hours of standing was easily distinguished from the effect of 9 hours of standing. This was another surprise. I had often stood 9 hours without noticing that I felt unusually refreshed the next morning; only when I managed to stand 10 hours did I notice this. Yet 9 hours and 10 hours of standing differ by only 10%. I knew of no psychology experiment in which a 10% change in duration had an easily noticed effect.

To find out if my impressions were correct, I began to rate how rested I felt when I awoke each morning. I used a 0–4 scale where 0 = completely tired, not rested at all and 4 = completely rested, not tired at all. A rating of 2 meant halfway between 0 and 4, a rating of 3 halfway between 2 and 4, and so forth. I made judgments to the nearest tenth (e.g., 3.2). I did not control how much I stood, except that I stood whenever it was not difficult to do so. I sat during meetings (because others were sitting); when eating alone, I sometimes stood.

Figure 13 shows results from the first six weeks of ratings. They confirmed my impressions: Rested ratings increased substantially as standing increased from 8 hours to 10 hours (upper left panel). The correlation (0.82) was reliable, $t(39) = 8.93$, $p < 0.001$. The ratings did not vary systematically with three other sleep parameters (the other three panels). None of the three corresponding correlations was reliable: time to bed and rested, 0.16, $t(39) = 1.03$; time out of bed and rested, -0.02, $t(39) = 0.11$; sleep duration and rested, -0.12, $t(39) = 0.74$.

To reduce the possible influence of expectations, I made a longer series of measurements, lasting 9 months, during which I tried to be as "blind" (ignorant of the treatment) as possible. Experiments could not be blind, so I continued to rely on outside forces to vary how much I stood each day. As mentioned earlier, I used the stopwatch on my wristwatch to measure how long I stood. I wrote down the accumulated duration many times during the day, zeroing the stopwatch each time, but I did not add up the collection of 10 or so durations at the end of the day. As a result, each evening I had only a vague idea of how long I had stood that day. I summed the durations for each day only after the study was finished. The results (Fig. 14, top left panel) – a correlation of 0.52, $t(192) = 8.52$, $p < 0.001$ – were essentially the same as the earlier results (Fig. 13, upper left panel), suggesting that exact knowledge of how long one has been standing has little effect.

Did something correlated with standing caused the change in rested ratings? The remaining five panels of Figure 14 help answer that question. If standing itself changed the ratings, the correlation would be "one-directional": hours of standing (e.g., on Monday) would correlate with a *later* rating (Tuesday morning) but not an earlier one (Monday morning). But if the correlation between standing and ratings was *not* due to causality, it could go both forward and backward in time. For example, if something caused both standing and rested ratings to be high during some weeks but not others, this would produce correlations in both directions; that is, standing would correlate with both earlier and later ratings. The top right panel of Figure 14 shows that there was no reliable correlation between rested ratings and standing the same day (e.g., rested ratings Monday morning and standing on Monday), $r = -0.06$, $t(189) = -0.87$.

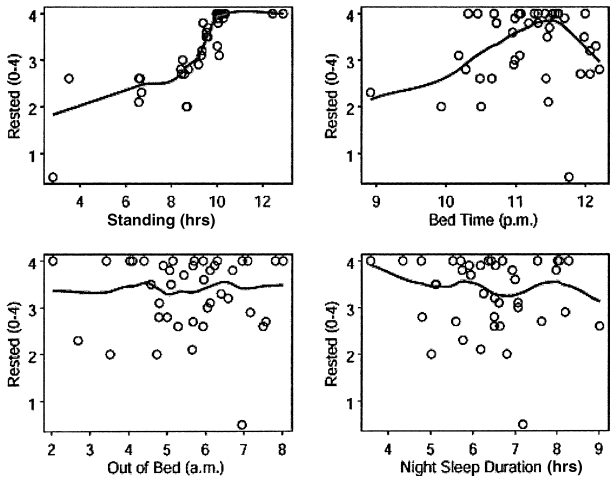

Figure 13. Relation between rested ratings (0 = completely tired, not rested at all; 4 = completely rested, not tired at all) and other measurements. Upper left: Correlation with duration of standing the previous day. Upper right: Correlation with bedtime. Lower left: Correlation with time out of bed. Lower right: Correlation with time slept. Data are from March 17, 1997 through April 28, 1997. Lines are based on loess.

The lack of correlation between rested ratings and same-day standing suggests that the search for important confounding variables should be restricted to the day before the rating (e.g., Monday, if the rating is Tuesday morning), the same period of time that standing is measured. Three measures of behavior during that period reliably correlated with standing: time spent sitting (-0.58, $t(188) = 9.68$, $p < 0.001$), time spent walking (0.56, $t(201) = 9.48$, $p < 0.001$), and total time awake (0.48, $t(188) = 7.48$, $p < 0.001$). (The middle left panel of Figure 14 shows the sitting/rested correlation.) All three correlated with rested ratings, but the sizes of the correlations (-0.26, 0.35, and 0.29, respectively) were less than the size of the standing/rested correlation (0.52). The absolute values of residuals from a loess fit of rested~standing (rested predicted from standing) were reliably less than the absolute values of residuals from a loess fit of (a) rested~sitting (paired comparison $t(185) = 4.48$, $p < 0.001$), (b) rested~walking (paired comparison $t(194) = 4.33$, $p < 0.001$), and (c) rested~time awake (paired comparison $t(185) = 4.04$, $p < 0.001$). (Degrees of freedom vary because of missing data.) In other words, the rested rating on Monday morning was better predicted by the duration of standing on Sunday than by the duration of sitting, walking, or time awake on Sunday. Figure 14 also shows the lack of correlation between the rested ratings and

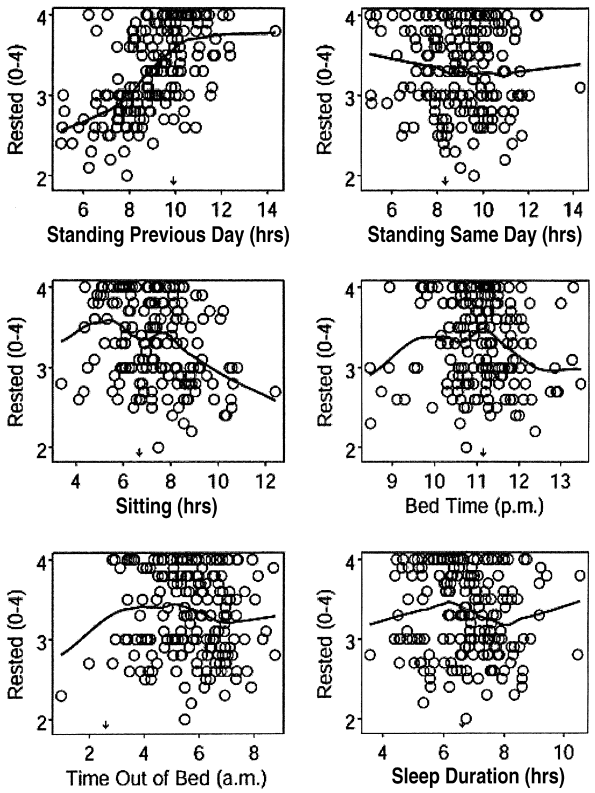

Figure 14. Relation between rested ratings (0 = completely tired, not rested at all; 4 = completely rested, not tired at all) and other measurements. Top left: Correlation with duration of standing the previous day. Top right: Correlation with duration of standing the next day. Middle left: Correlation with duration of sitting the previous day. Middle right: Correlation with time to bed (time of lights off). Bottom left: Correlation with time of getting out of bed. Bottom right: Correlation with sleep duration. Data are from April 29, 1997 through January 22, 1998. Lines are based on loess. The arrow in each panel indicates the position on the abscissa of a datum (rested rating = 1) outside the range of the ordinate.

time to bed (middle right panel), time out of bed (bottom left panel), and sleep duration (bottom right panel).

Another possible explanation of the standing/sleep correlation begins with the idea that the longer you are awake, the more tired you will be at bedtime; the more tired at bedtime, the longer you sleep; the longer you sleep, the more rested you will feel when you awake. Standing duration was positively correlated with total time awake, giving this possibility some credence. Figure 15 shows the relationship between rested ratings and standing, controlling for time awake and sleep duration. The data were divided into four groups depending on whether time awake and sleep duration were above or below their medians. The correlation between rested ratings and standing is clear and roughly the same in all four graphs. All four correlations were reliable, with all $t > 3.08$, two-tailed $p < 0.004$.

A commonsense explanation of the effects of standing is that standing is tiring; and the more tired you are the better you sleep. Indeed, I found that standing many hours each day was tiring the first few days. After that, however, it was not noticeably tiring. (The only obvious effect of standing a lot was that after about 10 hours, my feet might hurt slightly.) The data allow two tests of this idea. One uses a measure of sleep latency: how quickly I fell asleep. If standing made me more tired than usual, it should have

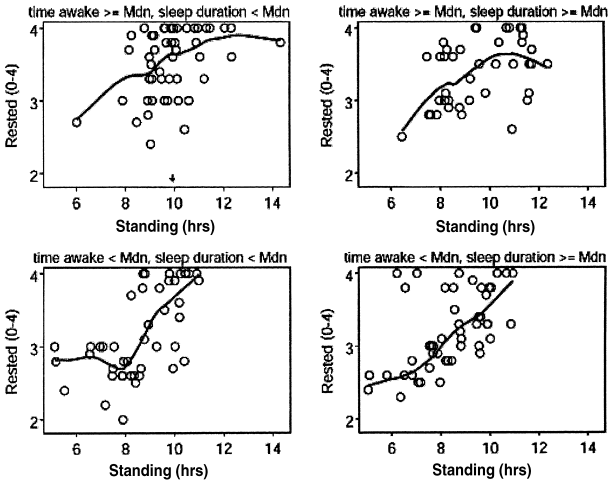

Figure 15. Effect of standing on rested ratings (0 = completely tired, not rested at all; 4 = completely rested, not tired at all) and standing, conditional on time awake and sleep duration. The data were split according to whether time awake and sleep duration were above or below their medians (4 categories). The arrow in the upper left panel indicates the position on the abscissa of a datum (rested rating = 1) outside the range of the ordinate. Data are from April 29, 1997 through January 22, 1998. Lines are based on loess.

made me fall asleep more quickly. As described in the Appendix, a measure of sleep latency was how many times I pressed a button (indicating I was still awake) in response to a beep that occurred 5 minutes after I started trying to go to sleep, and which recurred every 5 minutes so long as I kept pressing the button. Assuming that a failure to press the button during the 4 minutes after a beep meant I was asleep, a count of zero indicated a sleep latency of 0–5 min; a count of one, 5–10 min; a count of two, 10–15 min; and so forth. To determine how standing affected the count, which was usually zero, I divided standing duration into tertiles (lowest third, middle third, highest third) and computed the average count for each tertile (Fig. 16, upper panel). Standing had no detectable effect on sleep latency, according to a one-way ANOVA, $F(2, 232) = 0.67$, $p = 0.5$. This was meaningful because a similar measure *was* affected by standing. The same device also counted how many times I pressed the button after falling asleep, a measure of how often I awoke during the night. The number of awakenings was influenced by standing tertile, $F(2, 232) = 6.09$, $p = 0.003$. When I stood at least 8.4 hours, I woke up less often than when I stood less than 8.4 hours (Fig. 16, upper panel).

If standing more than usual made me more tired than usual, it should have made me sleep more than usual. Contrary to this prediction, there was a reliable *negative* correlation (−18) between standing and sleep duration, $t(237) = 2.78$, $p = 0.006$. As standing increased from 8 hours to 10 hours, sleep duration decreased from about 8.0 hours to 7.5 hours (Fig. 16, lower panel).

2.4.4. Related results. No prior sleep research hints at the importance of standing unless one considers it meaningful that astronauts have slept badly (e.g., Gundel et al. 1997). However, two observations suggest that our bodies are designed to stand much more than Americans typically do. The first is *the high rate of osteoporosis* (low bone density). About 20% of American women over 50 suffer from osteoporosis and many more suffer from osteopenia, a milder

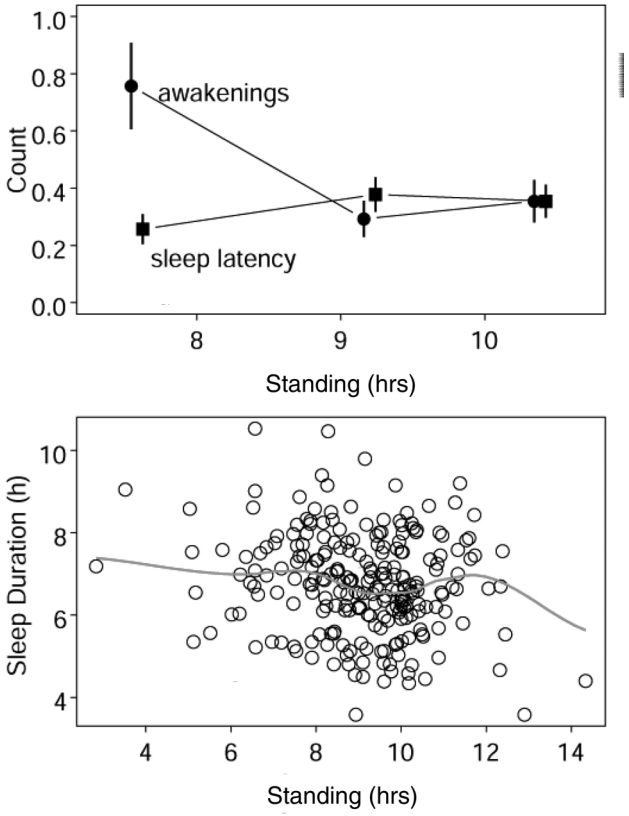

Figure 16. Effect of standing on sleep measures. Upper panel: Measures of sleep latency and nighttime awakenings versus standing tertile. Lower panel: Sleep duration versus standing duration. Sleep duration included episodes of falling back asleep after waking up. Data are from March 17, 1997 through January 22, 1998. Line is based on loess.

form of low bone density (Looker et al. 1997). Regular weight-bearing exercise is a standard treatment for osteoporosis and correlates with increased bone density (Smith 1985). The exercise need not be strenuous; modest amounts of walking increase bone density throughout the body (Krall & Dawson-Hughes 1994). The second is *the distribution of muscle fiber types.* Muscles consist of bundles of two types of muscle fibers: slow-twitch and fast-twitch. Slow-twitch fibers are capable of much longer use without rest than are fast-twitch fibers. Almost all muscles have about 40% slow-twitch fibers (Johnson et al. 1973). However, the proportion of slow-twitch fibers in the soleus (calf muscle) is about 90%, the highest in the body; the only other muscle with a similarly high proportion is a thumb muscle used for grasping.

2.4.5. Discussion. Sleep researchers take for granted that sleep varies in depth (e.g., the four stages of non-rapid-eye-movement sleep). It is assumed that the deeper the sleep, the more restorative it is and the harder it is to awaken the sleeper. More standing and deeper sleep (in this sense) were linked seven times in my experiments: (a) the reduction in early awakening from Baseline to Phase 1 (Fig. 12, upper panel); (b) the within-phase correlation between standing and rate of early awakening during Phase 1 (Fig. 12, lower panel); (c) the reduction in early awakening from Phase 1 to Phase 2 (Fig. 12, upper panel); (d) the within-phase correlation between standing and rate of early awakening during Phase 2 (Fig. 12, lower panel); (e) the first correlation between standing and rested ratings (Fig. 13, upper left panel); (f) the second correlation between standing and rested ratings (Fig. 14, top left panel); and (g) the reduction in awakenings with increased standing (Fig. 16, upper panel).

The simplest explanation of these findings is that standing made sleep deeper. Several other explanations are inconsistent with one or more results. At least some of the results cannot be due to expectations because the first experimental effect (the change from Baseline to Phase 1) and the first correlational effect (during Phase 1) were unexpected. The effect of standing on rested ratings (Figs. 13 and 14) was also surprising, at least at first. Although some of the instances are correlational (with the possible explanations that that opens up), some were experimental – that is, the change in standing was deliberate (the between-phase differences; upper panel of Fig. 12). Some of the correlational effects (Figs. 13 and 14, upper left panels) cannot be explained by correlations between standing and other sleep parameters (Figs. 13–15). That the rested ratings correlated with the previous day's standing (Fig. 14, top left panel) but not with same-day standing (Fig. 14, top right panel) is inconsistent with many noncausal explanations. Although duration of standing is almost inevitably correlated with the duration of other activities (e.g., time awake), in three cases (duration sitting, duration walking, and time awake) the correlated measure predicted the rested ratings more poorly than duration standing. With time awake controlled, the relation between standing and rested ratings remained (Fig. 15).

The longer we go without sleep, the more tired we become. Theories of sleep assume that this is due to something (often called Factor S) that increases during wakefulness and decreases during sleep (e.g., Borbely & Achermann 1992; Folkard & Akerstedt 1989; Gillberg 1997). The more Factor S, the more tired you feel. To understand two ways that standing might influence sleep, think of sleep as fire and Factor S as wood. Sleep burns up Factor S; the deeper the sleep, the hotter the fire. The more of Factor S that is "unburnt" when you awake (when the fire goes out) the more tired you feel. One possibility is that standing affects sleep by increasing the build-up of Factor S. More standing causes more Factor S to accumulate by bedtime (like having more wood on the fire when it starts). But this is almost the same as the commonsense notion, considered earlier, that standing is tiring, which is inconsistent with several results. Another problem with this possibility is that standing increased rather than decreased rested ratings (Figs. 13 and 14). If more Factor S is present when you go to sleep, more (or at least not less) should be present when you wake up, given the well-accepted idea that the timing of sleep is partly controlled (gated) by a circadian oscillator that tends to make us awake during the day and asleep at night (e.g., Borbely & Achermann 1992; Folkard & Akerstedt 1989; Gillberg 1997). This oscillator limits sleep duration.

Consistent with all of these data is that standing acts on sleep like an accelerant. Accelerants, such as gasoline, cause wood-fueled fires to burn hotter and faster. A gasoline-soaked piece of paper burns in a flash; without gasoline it burns more slowly. By deepening sleep, standing makes Factor S dissipate more quickly. Because Factor S is part of what causes sleep, this shortens sleep (Fig. 16, lower panel). To explain the effects of standing on early awakening (Fig.

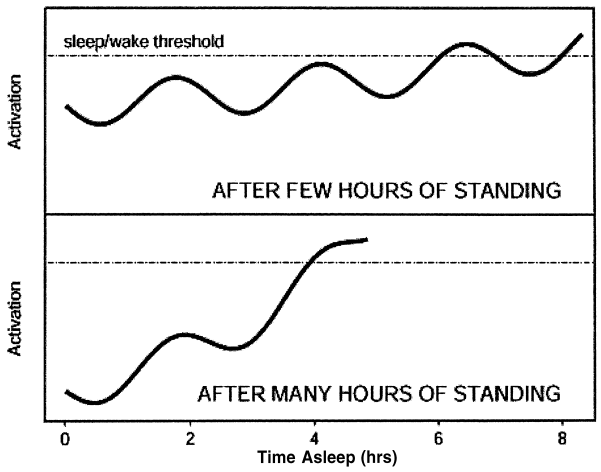

Figure 17. Explanation of the effect of standing on early awakening and sleep duration. Both panels show how activation, due to the depletion of a sleep-causing substance, increased during the night. After only a few hours of standing (upper panel), activation was higher at the start of sleep but increased more slowly than after many (8 or more) hours of standing (lower panel).

12) and awakenings (Fig. 16, upper panel) requires an additional assumption: the well-accepted idea that there is an ultradian (e.g., a 90-min) cycle of alertness (e.g., Orr et al. 1974). With little standing, activation stays close to the "surface" (the wake threshold) during sleep, making it more likely that the ultradian rhythm will be enough to produce wakefulness. Figure 17 illustrates this. To explain the effect of standing on rested ratings (upper left panels of Figs. 13 and 14) requires a different added assumption: the idea, mentioned in the preceding paragraph, that the timing of sleep is gated by a circadian oscillator. Overall sleep tendency depends on both the level of Factor S and the circadian variation, but how rested you feel depends only on Factor S. Because it wakes us in the morning, the circadian oscillator limits sleep duration. If a piece of wood can burn for no more than 1 minute (after 1 min, the fire is put out), soaking it with gasoline ensures that less wood remains when the fire ends.

What prehistoric problem did a standing-sleep connection solve? One plausible answer is the need to be on one's feet all day, for instance to forage or travel. Leg muscles need maintenance, which presumably cannot be done while they are active. Sleep can be used to perform maintenance on the muscles that are in continuous use when a person stands all day. But once these muscles are shaped to *use* sleep for maintenance, they will *need* sleep for that purpose. When the standing-sleep connection evolved, let us assume humans stood between 8 and 16 hours most days. At the low end of this range (8 hrs), leg muscles got many hours of rest during waking hours, and a typical amount of sleep provided enough time for the remaining maintenance that was needed. But when a person stood all day, maintenance had to be done entirely at night. A typical amount of sleep was not enough, so a mechanism that increased sleep duration after all-day standing was useful. Let us assume it was easier to build such a mechanism if it was allowed to produce *some* pressure to sleep after 8 hours of standing. So the mechanism that lengthens sleep after large amounts of standing produces some pressure to sleep after only 8 hours of standing, and the overall sleep regulatory system assumed that standing would provide a certain minimum amount of pressure to sleep. Without this minimal pressure to sleep provided by minimal standing (8 hrs would be minimal by Stone-Age standards), sleep is too shallow. This argument, although full of hard-to-check assumptions, does make a prediction: As standing duration increases from 8 hours to 16 hours, sleep duration should increase. Over a lower range, standing *decreased* sleep duration (Fig. 16, lower panel), so this prediction could plausibly be wrong. An alternative evolutionary explanation (suggested by Peter Totterdell) is that in the Stone Age the duration of standing acted as a measure of work to be done. The more work to be done, the more standing, and therefore the greater the need for fast, restorative sleep. This explanation predicts that as standing duration increases from 8 hours to 16 hours, sleep duration should decrease still further, or at least not increase.

Self-experimentation allowed me to test an unlikely idea (that standing causes weight loss), try a difficult treatment (standing 8 hrs/day for many days), and detect a change on a dimension far from the initial focus of interest (weight). It found a connection that previous research had not even hinted at.

I eventually got rid of the treadmill in my office, replacing it with a standing-level desk and an Orthomat (a mat designed for jobs that require standing).

2.5. Example 4: Morning light → Better sleep

2.5.1. Background. In 1997, I suggested to a depressed friend that he get more morning light, which has been used to treat depression (e.g., Wehr 1990a). He did so and felt a little better. I wondered if morning light would raise *my* mood.

2.5.2. Birth of the idea. To find out, I took morning walks during the summer of 1997. Over five weeks, I took 13 morning walks, irregularly spaced. They started at 6:00 a.m. and lasted a median of 60 minutes (range 30–80 min). My mood did not noticeably change. After a few walks, however, I noticed that the next morning I felt more rested than usual. For the 20 days before morning walks, the next-morning rested ratings had a median of 3.2 (interquartile range 3.0–3.6); after the first four walks, the next-morning ratings were 3.8, 4.0, 3.8, and 3.8. If these four ratings were independent, the probability that they were all outside the interquartile range in the same direction due to chance is 0.008.

The effect was not due to more standing because the walks (outdoors) had replaced walking on an indoor treadmill. When I noticed the effect, however, I did not yet know how long I had stood each day because I had not yet added up the many small times for each day to get daily totals. Figure 18 shows an analysis that takes standing into account. It compares ratings after 12 walks (one walk day was omitted because the standing duration was not available) with ratings from three sets of baseline days: (a) 50 days before the first walk, (b) days without walks between the first and last walking days, and (c) 50 days after the last walk. There was no systematic difference between the three sets of days, so the fitted line in Figure 18 is based on all of them. About half of the baseline points are above the line. However, 11 of the 12 walk ratings are above the line, reliably more than half (sign test, two-tailed $p = 0.006$). Because of the well-

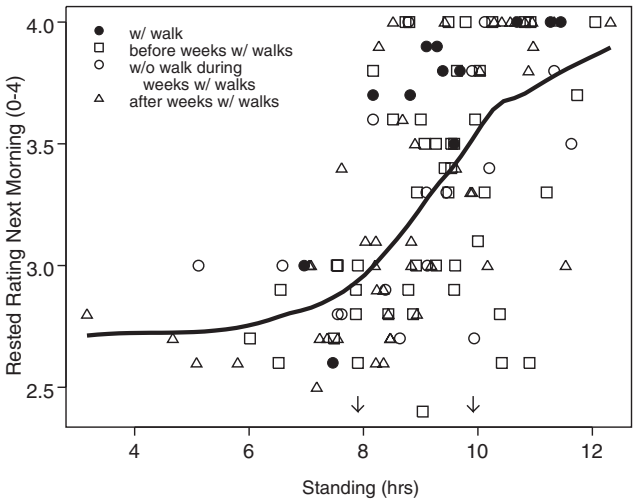

Figure 18. Effect of morning light on rested ratings. The line was fitted by loess to the baseline days. Data are from April 27, 1997 to October 22, 1997. Arrows indicate out-of-range points (rested ratings less than 2.4 on a scale where 0 = completely tired, not rested at all; 4 = completely rested, not tired at all).

known effect of light on circadian rhythms, exposure to sunlight was probably the crucial feature of the walks.

2.5.3. Test. To test the idea that exposure to light was the crucial feature, I did an ABABA experiment, with A = added morning light and B = no added morning light. The experiment took place indoors. The light came from an ordinary fluorescent fixture with four fluorescent lamps covered by a diffuser. (See the Appendix for more information.) On *light-on* days the lamps were turned on at 7:00 a.m. and turned off when I finished watching television, a median of 2.3 hours later (range 1.6–2.6 hrs). On *light-off* days the lamps were off (light from a nearby window remained). Nothing else was varied between the two sets of days. During the experiment I avoided any activities that would cause me to go to sleep later than usual, but I did not restrict my activities in other ways. To measure how rested I felt in the morning, I wrote down how much longer I wanted to sleep when I woke up. Such ratings are easier to understand than the rested ratings used in Example 3, which I also made.

The main results are shown in the top panel of Figure 19. Sleep-wanted ratings were always zero after light on, but never zero after light off, sign test, $p < 0.001$ (top panel). (Rested ratings showed the same effect.) This was not due to differences in standing durations because those were roughly the same in the two conditions (Fig. 19, middle panel). The bottom panel of Figure 19 shows when the lamps were on. The light had no clear effect on when I went to bed, how long I slept, or when I got out of bed (Fig. 20).

The light from the lamps changed my mood (Fig. 21), which I had not noticed at the time. (Perhaps the change was too small to notice without a special effort.) The data in Figure 21 are fragmentary because they do not include data from days after I had face-to-face social contact after 7 p.m. and because on several days I forgot to make the ratings (I did not expect them to matter). The direction of the effect depended on the time of day that the ratings were made: The light increased mood at 11 a.m. by 3 points, $t(14) = 3.72$, two-tailed $p = 0.002$. It increased mood at 4 p.m. by 5 points, $t(17) = 6.87$, $p < 0.001$. In contrast, it lowered

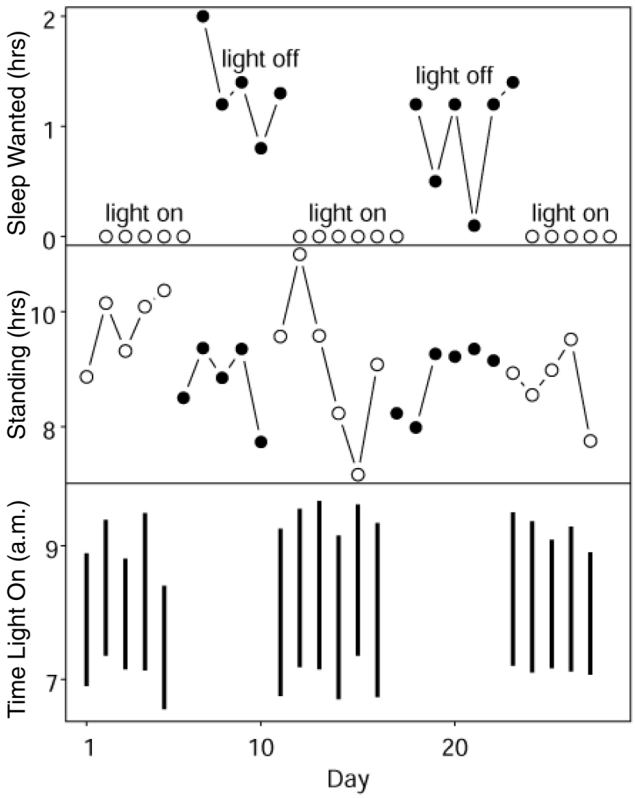

Figure 19. Effect of artificial morning light on sleep-wanted ratings the next morning. Top panel: How much more sleep I wanted when I awoke. Middle panel: Duration of standing. Bottom panel: Bars indicate when the artificial light was on. Day 1 = April 10, 1998. Day 1 was a Friday. There is no sleep-wanted rating (top panel) for Day 1. The sleep-wanted rating for Day 2 is from Saturday morning. The standing duration (middle panel) for Day 1 is from Friday. The timing of artificial light information (bottom panel) for Day 1 is from Friday morning.

mood at 10 p.m. by 7 points, $t(18) = 5.03$, $p < 0.001$. It also lowered mood at 7 a.m. The 7 a.m. ratings were made before the lamps were turned on (on light-on days), so the only comparison that makes sense is between the 7 a.m. ratings on the days after light on and the 7 a.m. ratings on the days after light off. For example, if Monday and Tuesday were light-on days and Wednesday and Thursday were light-off days, the comparison would be between Tuesday and Wednesday ratings and Thursday and Friday ratings. This comparison showed that the light lowered mood ratings by 4 points, $t(20) = 3.12$, $p = 0.005$.

After this experiment, I realized that the main result, that morning light made me wake up more rested (Fig. 19, top panel), resembled previous observations. As mentioned in the introduction to Example 1 (sect. 2.2.1), my early awakening had started around the time I put fluorescent lamps in my bedroom, timed to turn on early in the morning. This co-occurrence had led me to assume that light from the lamps caused early awakening. As stated earlier, I doubted that something as natural as sunrise would cause something as dysfunctional as early awakening, so I concluded that the early awakening was probably due to a difference between the artificial light and sunrise.

During sunrise, light intensity increases gradually, whereas the lamps in my bedroom produced an intensity-versus-time step function with only four levels, darkness and three levels of light. The light fixture in my bedroom

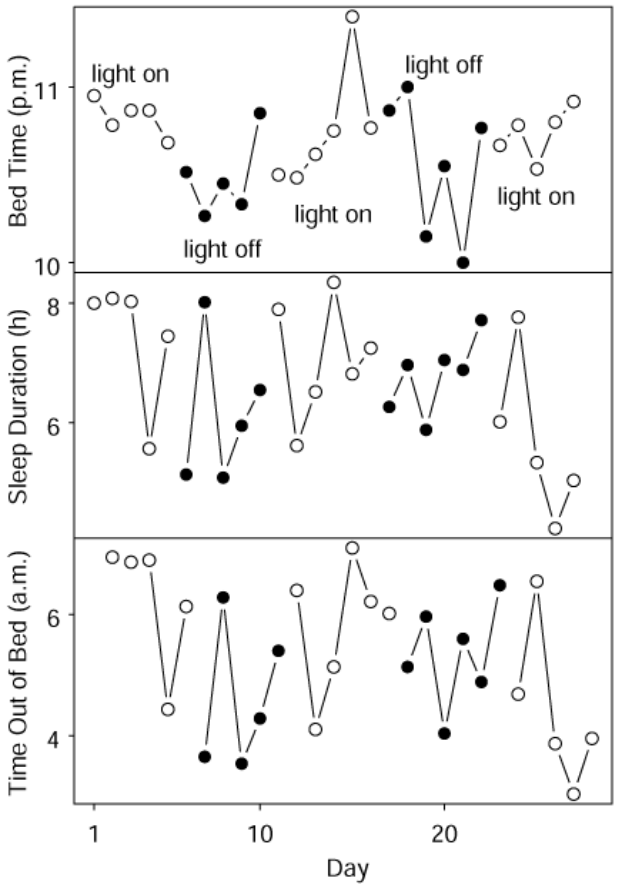

Figure 20. Effect of artificial morning light on sleep parameters the subsequent night and morning. Top panel: Time to bed. Middle panel: Sleep duration. Bottom panel: Time out of bed. Day 1 = April 10, 1998. Day 1 was a Friday. The time to bed (top panel) for Day 1 is from Friday evening. The sleep duration (middle panel) for Day 1 is from Friday evening-Saturday morning. There is no time out of bed (bottom) for Day 1. The time out of bed for Day 2 is from Saturday morning.

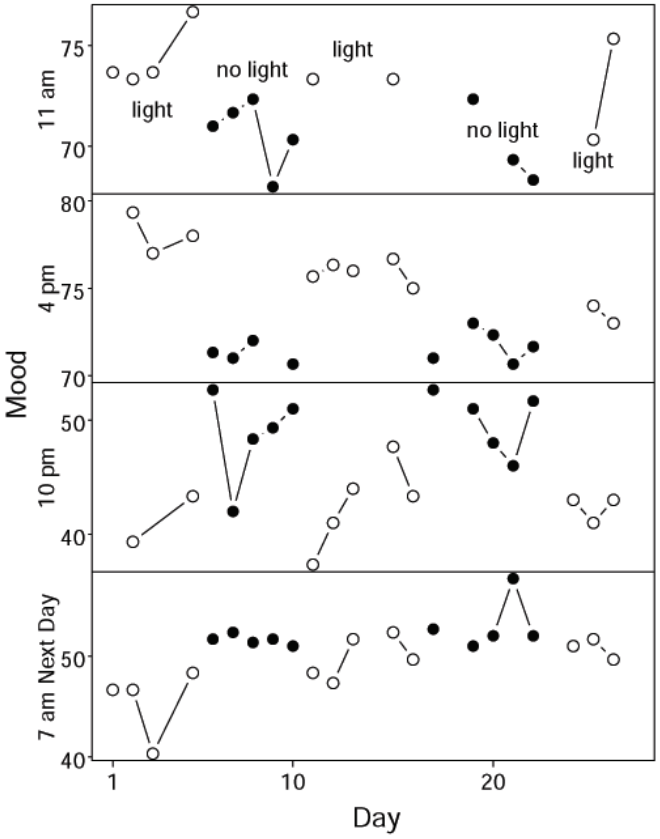

Figure 21. Effect of artificial morning light on mood. The bottom panel shows ratings of mood at 7 a.m. the next day – the morning after the light, if any. Day 1 = April 10, 1998. Each point is the average of three ratings, one per scale. The three scales measured dimensions unhappy/happy, irritable/serene, and reluctant/eager (scale range: 5 = extremely negative to 95 = extremely positive, with 50 = neither negative nor positive).

was on the floor and light from its lamps did not reach my eyes directly. I placed a light meter at the position of my eyes, pointed it in the direction of the lamps, and measured the three intensities of light that would reach me as 3, 13, and 57 lux. It was hard to smooth the intensity-versus-time function during a single day but the average over days could easily be smoothed out by starting the lamps at a different time each morning (e.g., at 5:10 a.m. one morning, at 5:40 a.m. the next morning, and so on, at randomly chosen times).

I tried this but it failed to reduce early awakening. I began to question my assumption – the basis of years of self-experimentation – that light from the lamps caused early awakening. As a test, I did an ABA experiment, with A = light on and B = light off. The leftmost three points in Figure 22 show the results, which, to my great surprise, implied that the light *reduced* early awakening. Comparing the proportion of days with early awakening during Conditions 1 and 3 (light on) together to Condition 2 (light off), I observed a reliable difference, χ^2 [1] = 12.61, $p < 0.001$.

These results were so unexpected that I repeated the test. The result was the same. Comparing Conditions 3 and 5 (light on) to Condition 4 (light off), I found a reliable difference, χ^2 [1] = 5.65, one-tailed $p = 0.01$. Still skeptical, I did the test a third time and again got the same result. Comparing Conditions 5 and 7 (light on) to Condition 6 (light off), I found a reliable difference, χ^2 [1] = 6.50, one-tailed $p < 0.01$. Because my ideas about the effect of the light on early awakening were clearly wrong, there was no longer any reason to randomly vary the starting time of the lights, and I stopped varying it. Then I tested the effect of the light again. The result was essentially the same. Comparing Conditions 8 and 10 (light on) to Condition 9 (light off), I found there was almost a reliable difference, χ^2 [1] = 2.33, one-tailed $p = 0.06$.

2.5.4. Related results. Thousands of experiments have shown that light affects the time of day that animals are active and several have shown that it can control when humans sleep (e.g., Czeisler 1995; Minors & Waterhouse 1986). Kohsaka et al. (1999) found that exposure to 1000 lux for 1 hour in the morning (time unspecified) improved a rating of sleep maintenance from the "OSA inventory" (Kohsaka et al., p. 240) of healthy elderly women. Kim and Tokura (1998) found that exposure to bright light from 7:30 a.m. to 6:00 p.m. increased the amplitude of the circadian body temperature rhythm; it became lower at night.

Bright light has been useful in the treatment of winter depression, but has failed to substantially reduce nonseasonal depression (Wehr 1990a; Wirz-Justice 1986). However, Leppämäki et al. (2002) found that bright light expo-

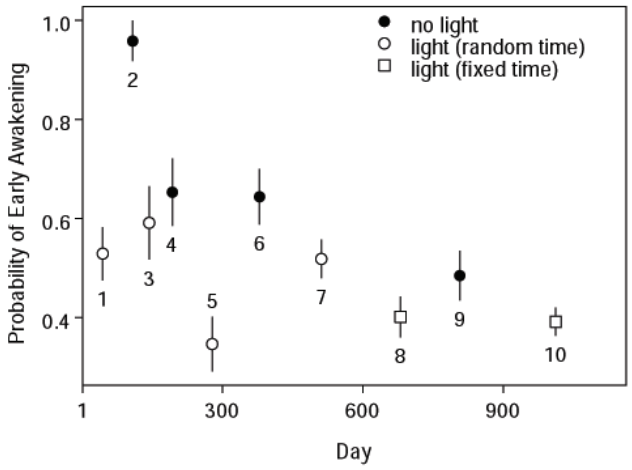

Figure 22. Effect of artificial morning light on the probability of early awakening. Day 1 = April 10, 1990. The number of days in Conditions 1–10 was 87, 24, 44, 49, 75, 73, 164, 142, 99, and 291, respectively. Vertical bars indicate standard errors.

sure improved the mood of healthy subjects more than exposure to light of ordinary indoor intensity.

2.5.5. Discussion. The two sets of observations (Figs. 18 and 19) make a good case that morning light made me feel more rested when I awoke. Though the later results (Fig. 19) could have been influenced by expectations, the initial discovery (Fig. 18) could not have been, and the earliest results (Fig. 22) were repeatedly the opposite of what I expected. Increasing sleep duration can make sleep more restorative (8 hrs of sleep will be more restorative than 4 hrs), but the light did *not* clearly change sleep duration (Fig. 20). Apparently the light made sleep deeper and therefore more efficient.

Given that light affects the timing of sleep, the *phase* of the sleep/wake rhythm, it is hardly surprising that it can also change the depth of sleep, the *amplitude* of the rhythm. Nevertheless, in spite of its commonsense nature, the notion that morning light can deepen sleep surprised me (more than once) and will be news to sleep researchers. In contrast to the thousands of studies about the effect of light on the phase of circadian rhythms, very few have measured its effect on amplitude, and the only ones I know of (Jewett et al. 1991; Winfree 1973) used light to reduce amplitude, rather than to increase it. One indication of the neglect of amplitude changes is that the term *phase response curve* (showing the change in phase of a circadian rhythm as a function of time of day of light exposure) has appeared in print countless times, whereas *amplitude response curve* has apparently never appeared in print. In a 76-page consensus report on the use of light to treat sleep disorders (the first chapter is Campbell et al. 1995), the possibility that amplitude changes might be helpful is never mentioned, in spite of the statement that "it is well accepted that bright light exposure can influence dramatically both the amplitude and phase of human circadian rhythms" (Campbell et al., p. 108). All of the beneficial effects mentioned are phase shifts. Yet the number of people who sleep at the wrong time (e.g., due to jet lag or shift work) is undoubtedly far less than number of people who would benefit from deeper sleep.

That early awakening – at least, mine – was an amplitude problem rather than a phase problem did not surprise me. The observation that standing much more reduced early awakening (Example 3), had implied this conclusion; because it was implausible that standing more, in bits and pieces over the course of day, could cause a phase change. However, early awakening has been seen as a phase problem by circadian-rhythm researchers. For example, Lack et al. (1996) concluded that "early morning awakening insomnia arises from phase advanced circadian rhythms" (p. 211). Based on this view, Allen (1991), Czeisler et al. (1989), and Lack and Wright (1993) used *evening* bright light to treat early awakening. Subjects woke up later, but whether they felt more rested when they awoke is unclear.

The mood changes produced by light (Fig. 21) were too small to be clinically useful, so they are consistent with the view that phototherapy is not helpful for non-seasonal depression. They have scientific interest, though, because they were bidirectional (better mood in the middle of the day, worse mood earlier and later). This pattern suggests that the light increased the amplitude of a circadian oscillator that controls mood, which is more support for the conclusion of Example 2 that such an oscillator exists.

2.6. Example 5: Standing and morning light → No colds

2.6.1. Background. American adults average 2 to 3 colds (upper respiratory tract infections) per year (Turner 1997). Research about the prevention of colds has tested drugs (Jefferson & Tyrrell 2001) and nutritional supplements (e.g., Josling 2001).

2.6.2. Birth of the idea. In the spring of 1997, I realized that during the winter I had not had any colds. The people around me had often been sick. The lack of colds continued till at least November 2002 (when this was written). Due to my sleep research, I had recorded every cold I caught since 1989. The top panel of Figure 23 shows the date and duration of my colds since then. The many connections between sleep and the immune system (see sect. 2.6.3) suggest that the decrease in colds began near January 16, 1997, the day I started to stand about 10 hours per day and as a result started to sleep very well every night (Example 3). Beginning in January 1998, I was exposed to about 1 hour of sunlight or artificial sunlight each morning, which reduced how much time I needed to stand to sleep well (Example 4). The lower two panels of Figure 23 show how much I stood each day and when I was exposed to morning light. Standing, as mentioned earlier, included both moving and standing still. The morning light, as mentioned earlier, was produced by four fluorescent light bulbs shining upward on my face as I watched TV while walking on a treadmill.

Is the decrease in colds reliable? The records include 2,518 days before January 16, 1997, and 1,799 days after that date, a total of 4,317 days. The records omit 409 days away from home, when conditions differed in many ways from home life. Over the whole recording interval I caught 19 colds. If the likelihood of colds were the same before and after January 16, 1997, then the probability that all 19 colds would occur before that date, which is what happened, is $(2518/4317)^{19} = 0.00004$. Thus, the decrease is highly reliable. This analysis assumes that colds were equally likely all year, when in fact colds are more common in winter. My 19 colds were unequally distributed, with 13 during the colder half of the year, October–March, and 6 during

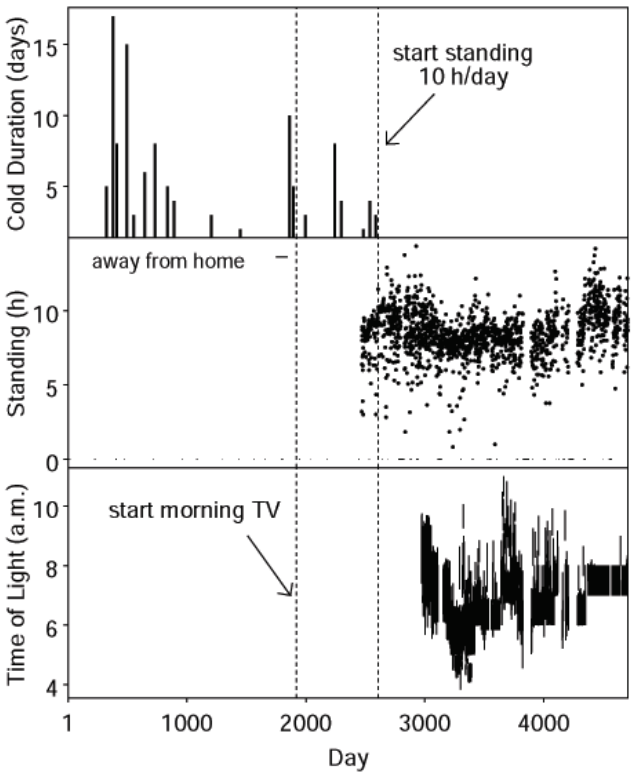

Figure 23. Health, standing, and morning light. Top panel: Incidence and duration of colds (upper respiratory tract infections). Middle panel: Duration of standing as a function of day. The lines at the bottom of the middle panel indicate days away from home. Bottom panel: Exposure to artificial morning light as a function of day. Day 1 = November 27, 1989, when record-keeping began. All times are Standard Time.

April–September. If the statistical test is adjusted so that days from October to March are given twice the chance of generating a cold as other days, $p = 0.00003$, essentially the same.

At the beginning of the cold-free period (January 1997 to November 2002) I noticed for the first time that sometimes I would be inexplicably tired for a day or two. It happened a few times per year. During these episodes I slept 2 to 4 hours per day more than usual and felt tired while awake. The extra sleep and fatigue were inexplicable in the sense that they had no obvious cause, such as poor sleep or exertion. In a few cases these bouts were preceded by a slightly "ticklish" throat, which before January 1977 had usually preceded full-blown colds. In all of these cases I may have been fighting off an infection. If so, it suggests that I was free of colds because my immune system did a better job of suppressing them. When persons are intentionally exposed to cold viruses, a large percentage of them do not catch colds (e.g., Cohen et al. 1997).

2.6.3. Related results. Sleep and the immune system are obviously connected. We sleep more when we are sick, and research has uncovered many other links (Benca & Quintans 1997; Dinges et al. 1995; Krueger & Karnowsky 1987; Moldofsky 1995). Especially revealing is the work of Pappenheimer and his colleagues (Pappenheimer 1983), who isolated a chemical in the cerebrospinal fluid of sleep-deprived dogs that when injected into other dogs made them sleepy. The chemical turned out to be a muramyl peptide. Muramyl peptides are bacterial fragments whose concentrations increase during a bacterial infection. This particular muramyl peptide was an immunostimulant and a pyrogen (Pappenheimer 1983). Viral fragments increase sleep as well (Krueger 1990), supporting the idea that my bouts of unusual sleepiness were due to fighting off cold viruses. Cohen et al. (1997) found that subjects who reported better sleep efficiency (percentage of time in bed spent sleeping) were less likely to come down with a cold when intentionally exposed to cold viruses.

2.6.4. Discussion. The decrease in colds was almost surely due to better sleep. My sleep dramatically improved (Examples 3 and 4) at about the same time I stopped catching colds. There were no other long-term changes in my life around that time (e.g., no changes in job, house, diet, or living conditions).

That better sleep can increase resistance to infection is not a new idea. The novelties here are (a) the size of the effect (complete elimination), which suggests that sleep that is sufficiently good can eliminate colds, and (b) what caused it (more standing and morning light). They suggest new explanations of some well-known facts:

1. *Infectious diseases are more common in winter* (e.g., Douglas et al. 1991). Upper respiratory tract infections are called *colds*, of course, because they are more common in winter. The new explanation is that the increase is due to less morning light. Example 4 suggests that a reduction in morning light will make sleep worse. In northern latitudes, such as Norway, it is well known that sleep gets worse during the winter (Hansen et al. 1998). Studying monthly rates of influenza around the world, Hope-Simpson (1981) found that the highest rates were associated with the sunlight minimum, not the temperature minimum, and concluded that the results "suggest a direct effect of variations in some component of solar radiation on virus or human host" (p. 35).

2. *Physical activity is associated with good health* (e.g., Blair et al. 1992). The earliest study usually cited to support the claim that physical activity is beneficial is a survey of London Transport employees working on buses and trams (Morris et al. 1953). The fare collectors (who stood almost all day) had much less heart disease than the drivers (who sat almost all day). Yet the fare-collectors did nothing that resembles current American conceptions of exercise (activities that produce a substantial and sustained increase in heart rate). Some of the fare collectors worked on double-decker buses with an open-air upper deck, so they would have experienced more morning light than the drivers, who were always inside. Morris et al. found similar differences between postmen and Post Office employees with sedentary jobs. The postmen were obviously outside more than the sedentary workers but whether this was early in the morning is not clear (mail is often sorted in the morning). Example 3 suggests that the workers who were on their feet all day slept much better than those who sat all day, and evidence is accumulating that infections contribute to heart disease (e.g., Danesh et al. 1997; Nieto 1998; Streblow et al. 2001; Weerasinghe et al. 2002). More recent epidemiological studies have emphasized that even moderate activity is associated with a large reduction in coronary mortality compared to being sedentary (Blair & Connelly 1996). The new explanation is that exercise is associated with better health (including less heart disease) because it increases standing and often increases exposure to morning light (e.g., walking or jogging before work), thereby improving

sleep and immune function. For instance, Nieman (1994) found that women who walked 40 to 45 minutes at least 5 days per week had half as many days with cold symptoms as women who walked much less. He did not measure exposure to morning light, which was probably confounded with walking – the more walking, the more morning light.

3. *Depression is associated with poor health* (Dinan 1999; Herbert & Cohen 1993; Vaillant 1998), including, at least in men, a much higher risk of heart disease (Dinan; Hippisley-Cox et al. 1998). Depression is also associated with poor immune function. Herbert and Cohen (1993, p. 481) noted that "depressed persons sleep less, exercise less, have poorer diets, smoke more, and use alcohol and other drugs more than do non-depressed persons," and of course there are other differences between the two groups. The present results, by suggesting that a change in sleep can have a big effect on immune function, focus attention on the very strong correlation between depression and insomnia (e.g., Ford & Kamerow 1989) as an explanation for the poor health and poor immune function associated with depression. The depression/poor health association may help explain why standing and morning light managed to *eliminate* colds, rather than just reduce them. In this example, standing and morning light may have improved sleep and immune function that were already good because I was watching faces on television almost every morning, which had effects opposite to depression (Example 2). Figure 23 suggests that my health did improve after I began watching faces on television most mornings (Day 1,919 in Fig. 23). From Days 1 to 1,918, I caught 13 colds (mean duration 6 days); from Days 1,919 to 2,607 (start of standing 10 hrs/day), 6 colds (mean duration 3 days). The difference in mean duration is reliable, with a 95% confidence interval of 0.2 to 5.9 (based on bootstrap, 10,000 replicates).

Americans miss many days of work because of colds. When Verifone, a manufacturer of key pads, added many skylights to one of its buildings to save energy, absenteeism decreased 45% (Browning 1997). Workplace changes that increase morning light and make it easier to stand while working might pay for themselves.

2.7. Discussion of Examples 1–5

"Evolutionary psychology is in many respects a strangely inconsequential exercise," wrote Horgan (1999, p. 194), because its adherents "cannot perform experiments that show that their view is right and the alternative view is wrong – or vice versa" (p. 194). In Examples 1–5, evolutionary ideas *were* consequential. An evolutionary interpretation of Example 1 led to Examples 2, 3, and 5, which included several experiments. The next section shows that evolutionary ideas were also helpful in the study of weight control. They provided initial support for a theory (see sect. 3.3.1, the Background section of Example 7) that eventually led to a very good way of losing weight (Example 10).

The mismatch idea – that many problems are due to differences between Stone Age and modern conditions – is easy to use: You point to Problem Y and say that it is (or, better, ask if it is) due to Absent-in-Stone-Age Element X. This may seem too easy to be helpful. For example, suppose I claim that obesity (Y) is caused by sugar consumption (X) (Steward et al. 1998). If I eliminated sugar from my diet and lost weight, would this support the mismatch idea? Yes, but only very slightly because almost our entire diet is non–Stone-Age. Why emphasize sugar? The connection between the mismatch notion and the specific recommendation is very weak. What about obesity (Y) and wearing socks (X)? This is not worth serious consideration. By itself, the mismatch idea is of little help because it points in too many directions. But when combined with (a) other ideas that narrow the possibilities (e.g., eliminate the socks idea), and (b) a cheap way to test the possibilities that remain, the mismatch idea can have considerable value, as Examples 2, 3, and 5 make clear.

Self-experimentation is not the only way to do this. Corruccini and Kaul (1983) measured rural and urban rates of dental malocclusion (e.g., crooked or crowded teeth), myopia, and chronic allergy and asthma in a genetically homogenous group of Punjabis living near Chandigarh, India. Much like the speculation behind Examples 2 and 3, they assumed that rural life was closer than urban life to the environment that shaped the human genome. For each disorder the rural rate was lower than the urban rate, which suggested possible causes. Using anthropological and quasi-experimental data, Corruccini (1999) persuasively argued that malocclusion is due to a too-soft diet, a very useful and non-obvious conclusion. According to the National Library of Medicine (2002), "malocclusion is most often hereditary" and "many types of malocclusion are not preventable." Just as it is hard to imagine conventional research discovering the effect of viewing faces in the morning on mood, or the effect of standing on sleep and colds, it is hard to imagine mainstream health research discovering the effect of diet on malocclusion. Standard epidemiology will not detect it because in America everyone eats a soft diet.

Most evolutionary psychology has used evolution as a source of explanations – a way to explain important facts such as male/female differences. If these explanations are untestable, this is, as Horgan (1999) pointed out, a serious weakness. Corruccini's work and Examples 2 and 3 illustrate a different use of evolutionary ideas – as a starting point. They combined the mismatch idea with basic anthropology to generate hypotheses worth testing. Example 2 used the conclusion that Stone-Age people lived in groups (anthropology at its most basic). Example 3 was partly based on the idea that Stone-Age people spent most of the day on their feet. In a way, anthropology and psychology need each other. Anthropology generates many ideas about human nature, but anthropologists have a hard time testing them. Psychologists do careful experiments and surveys, which are great for testing ideas but not so good for generating them.

3. Pavlovian weight control

3.1. Introduction

The set point theory of weight control (Kennedy 1950) assumes that adipose fat is regulated by a feedback system with a set point. Its core assumption is that when the amount of fat is less than the set point, changes occur (e.g., more hunger) that increase the amount of fat. This idea is widely accepted (e.g., Hirsch et al. 1998; Keesey & Hirvonen 1997; Michel & Cabanac 1999; Schwartz et al. 1999), with some dissent (e.g., Bolles 1990; Garrow & Stalley 1975; Wirtshafter & Davis 1977). The set point theory was the starting point for the ideas developed here.

In 1996, when Examples 6–10 began, most researchers believed (and still believe) that the set point is not affected

by the environment (Gibbs 1996; Kolata 2000). For instance, "although the set point may change with age, it does so according to a fixed genetic program; diet or exercise can move you away from your set point, at least for a time, but the target itself cannot change" (Gibbs, p. 91). Keesey and Hirvonen (1997) concluded that a half-year exposure to a high-fat diet could irreversibly raise the set point, but mentioned no other ways that diet can affect the set point. Michel Cabanac was the only person to take for granted environmental control. He and his colleagues had done brilliant work about how taste (Cabanac & Rabe 1976; Fantino 1976) and exercise (Cabanac & Morrissette 1992) affect the set point, but this had not been followed up.

Implicit in the set point theory is the assumption that it is hard to stay at a weight much less than your set point. If you lose substantial weight easily and stay at the lower weight, your set point must have decreased. Given this, the observation that eating less-processed food allowed me to lose weight without effort (Fig. 1, lower panel) and remain at the lower weight for years implied that the dietary change had lowered my set point. This was impressive confirmation of the Cabanac view, and it raised the question: By what rules does diet control set point? The processing observation was too complicated to help here. To induce the rules, I needed simpler dietary changes that affect the set point. Examples 6–10 describe how I found such changes and a theory that explains their effects. This section is called "Pavlovian weight control" because the theory (described in the introduction to Example 7) assumes that flavor-calorie associative learning is very important.

3.2. Example 6: Water → Weight loss

3.2.1. Background. It was not obvious what simple dietary changes might lower my set point (I did not want to gain weight). I already avoided the usual suspects, such as baked goods, candy, soft drinks, bread, and desserts. Because the search was difficult, I was willing to try dietary changes that could not be permanent.

3.2.2. Birth of the idea. In 1996, an acquaintance told me that when he lived in Japan he began eating a macrobiotic-like diet based on Diamond and Diamond (1985), a diet that included lots of fruit, vegetables, brown rice, and water, along with small amounts of fish and chicken (Tray Critelli, personal communications, June 5, 1996 and August 8, 2000). Starting at 86 kg and a body mass index (BMI, kg/m^2) of 25, he lost 20 kg, ending with a BMI of 19, which is quite thin. At the time my BMI was 26. Except for the water, his diet resembled mine. Did water make a big difference?

3.2.3. Test. I began drinking much more water than usual. Within a few days my water intake reached 5 liters/day (not counting water in food), close to the maximum I could tolerate. I had not measured my prior intake, but it had surely been less than 1 l/day. I did not measure my calorie intake nor consciously try to control it. I simply ate enough to feel comfortable. The set point theory came from rat experiments in which the rats ate ad libitum (as much or as little as desired).

I was already measuring my weight regularly. I used three scales, to ensure that any weight change was not due to a changing scale. Figure 24 shows the results. I lost about 2.8 kg over two weeks before my weight stabilized. After 4 months, tired of drinking so much water, I reduced my intake from 5 l/day to 3 l/day. My weight rose about 0.7 kg. In the middle of the 5 l/day period I changed my diet (see Example 7), which is why my stable weight at the end of the 5 l/day period (78.6 kg) was less than my stable weight near the start of that period (80.5 kg).

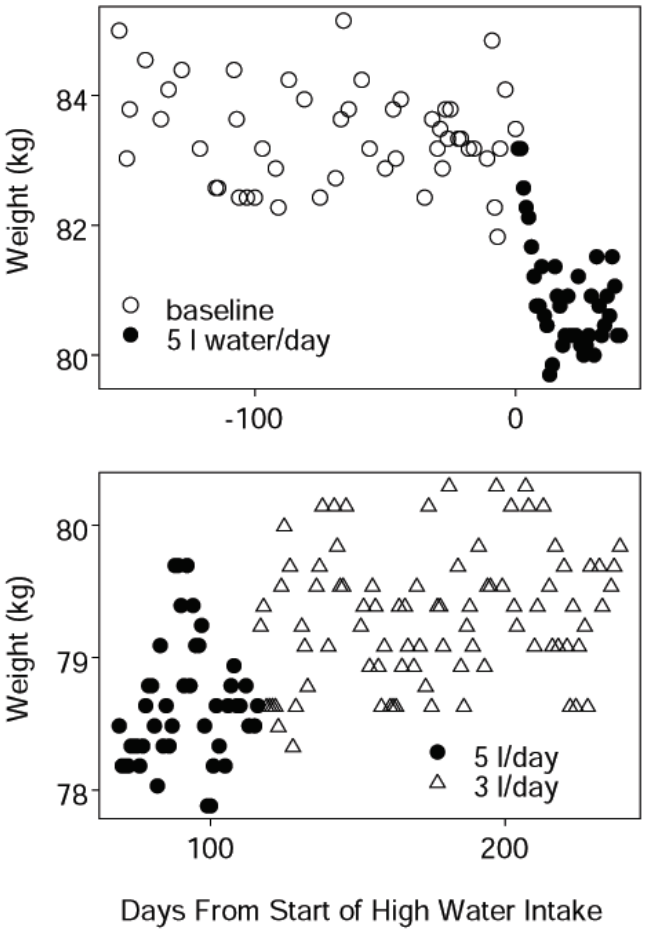

Figure 24. Effect of water intake on weight. High water intake began June 6, 1996.

3.2.4. Related results. There are no similar results in the scientific literature, as far as I know. Studies of the effect of water on weight have used relatively small amounts of water and measured how much is eaten at the next meal.

3.2.5. Discussion. The practical value of these results is slight because easy-to-drink amounts of water (1–2 l/day) had only a small effect. However, they show that self-experiments about weight control can have interesting results and they challenge mainstream ideas on the subject (e.g., Keesey & Hirvonen 1997), which emphasize amount and source of calories (e.g., percentage of calories from fat). Drinking water changes nothing that mainstream theories say is important.

The results are not easy to explain. That the water was "filling" and therefore I ate less at meals (Brownell 1987; Tennesen 2000) is unlikely because I drank almost all the water between meals. Apart from the difficulty of drinking so much water, it was easy to stay at the lower weight, suggesting that the water intake lowered my set point. But, at the time, no explanation of why it should do so came to mind.

Water consumption is often recommended for weight loss (e.g., Brownell 1987; Heller & Heller 1997; Irons 1998; Sears 1995; Tennesen 2000) but without evidence.

3.3. Example 7: Legumes → Weight loss

3.3.1. Background. After the weight-loss phase of Example 6, I happened to read Ramirez (1990a), who described experiments with rats that measured the effect of adding saccharin to a liquid diet. Rats that got saccharin weighed more than rats that did not. The commonsense explanation, that the sweetness of saccharin made the liquid diet taste better, was not consistent with the effect's long latency (it took more than a week to appear) and fragility (it was eliminated by prior exposure either to saccharin or to the liquid diet).

That the two groups (with and without saccharin) weighed the same after one week suggested that learning was involved. The pre-exposure effects pointed to a particular type of learning: Pavlovian conditioning. Many examples of Pavlovian conditioning show a *CS pre-exposure effect:* Exposure to the conditioned stimulus (CS) without exposure to the unconditioned stimulus (US) reduces the effect of later CS-US pairings (e.g., Lubow 1973). Also common is a *US pre-exposure effect:* Exposure to the US without the CS reduces the effect of later CS-US pairings (e.g., Randich & LoLordo 1979). That prior saccharin exposure reduced the effect of saccharin resembled a CS-pre-exposure effect. That prior exposure to the liquid diet reduced the effect of saccharin resembled a US-pre-exposure effect. The similarities suggested that an association between saccharin and the liquid diet was behind the saccharin-induced weight gain. The existence of flavor-calorie associations is well-established (Sclafani 1991), at least in rats. The saccharin would provide the flavor; the liquid diet, the calories.

Because the rats ate ad libitum, the weight difference implied a set point difference. So Ramirez's (1990a) results suggested to me that the set point was sensitive to flavor-calorie associations. The results were easy to explain if I assumed that saccharin caused weight gain only when its flavor was associated with the calories in the liquid diet. It was a short step from this to a general theory. The theory adds two assumptions to the set point idea: (1) *Flavors associated with calories raise the set point.* The stronger the association, the greater the increase. (2) *Between meals, the set point falls.* It falls more slowly than the actual level of adipose fat. The greater its value, the faster it falls. For example, the set point might fall 1 kg/day when at a high value and 0.5 kg/day when at a low value.

To illustrate these assumptions, Figure 25 compares the effects of a carrot and a cookie. The two foods have the same number of calories (we'll assume) so they increase weight the same amount. But because the flavor of the cookie is more strongly associated with calories than the flavor of the carrot, the cookie raises the set point more than the carrot. Figure 25 also illustrates the assumptions that (a) during periods without food, the set point falls more slowly than the actual level of fat, and (b) as the difference between the set point and the actual amount of fat increases, hunger increases (part of the original set point theory).

The set point will stabilize at the weight where its

decrease between meals = increase at meals.

The left side of the equation depends on your weight – the greater your weight, the greater the decrease between meals. The right side of the equation – the increase at meals – depends on what you eat. Everyone needs to consume a certain amount of energy per day to fuel activity and metabolism. How much that consumption raises the set point depends on the source of the energy – some foods raise the set point more than other foods with the same number of calories. When the two sides of the equation are not equal, your set point will gradually move up or down – pulling your weight along with it – until they become equal.

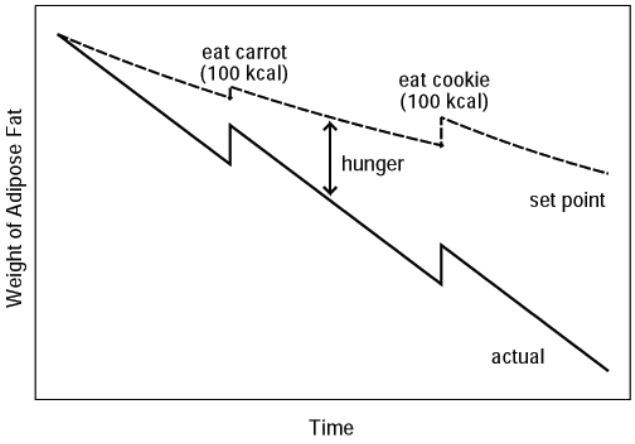

Figure 25. Assumptions of new weight-control theory: How foods with the same effect on body fat vary in their effect on the body-fat set point. The decelerating decline in the set point is part of the theory. That the decline in actual body fat is linear is not part of the theory.

Figure 26 shows why two foods with the same number of calories can raise the set point by different amounts. The strength of a flavor-calorie association depends on how much the flavor signal (the neural signal generated by a flavor) overlaps the calorie signal (the signal generated by the detection of calories). The more overlap, the stronger the association. This is surely true of any Pavlovian association: Its strength depends on the overlap in time and space between the neural events caused by the CS and the neural events caused by the US. Figure 26 indicates that there are

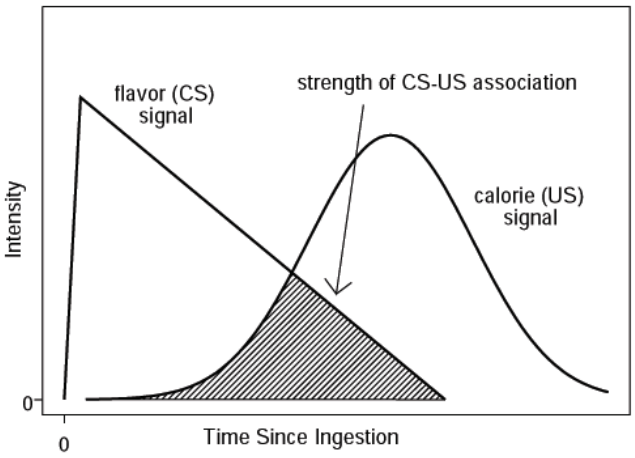

Figure 26. How taste and calorie signals determine the strength of taste-calorie associations. The strength of the association depends on the amount of overlap–the more overlap, the stronger the association. CS = conditioned stimulus; US = unconditioned stimulus.

two ways to reduce the strength of association while keeping calories constant: *reduce flavor,* thereby reducing the size of the flavor signal; and *delay detection of calories* by slowing down digestion. The calorie signal cannot just be lowered because that would correspond to eating fewer calories.

This theory gained immediate support, at least in my thinking, from two of its features. First, it made evolutionary sense. Adipose fat stores energy. The logic of storage systems, such as grain elevators, is that you stock up when the price is low and "stock down" when the price is high. So we should have more adipose fat when food is plentiful than when it is scarce. That is what the mechanism that this theory describes makes happen. When food is plentiful, you have more choice of food. The more choice, the stronger the strongest flavor-calorie associations of the available food will tend to be (just as the larger a group of people, the taller the tallest person in the group will tend to be). In rat experiments, flavor-calorie associations have been detected with preference tests: Foods with a calorie-associated flavor are preferred to foods without such a flavor. Given a choice of several foods, then, we tend to choose the one with the strongest flavor-calorie association. So the more choices we have, the stronger the flavor-calorie associations of what we choose. (Famine foods – foods of last resort – illustrate this. They have low energy density and produce very weak flavor-calorie associations.) The stronger the flavor-calorie associations of your diet, according to the new theory, the higher your set point. The higher your set point, the higher your weight.

Second, the theory easily explained the observation that eating less-processed food caused weight loss (Fig. 1, lower panel). The explanation is that processing usually increases the strength of flavor-calorie associations. This was quite plausible. As just mentioned, flavor-calorie associations have been detected with preference tests: the stronger the association, the stronger the preference. My dietary shift to less-processed foods (e.g., from orange juice to oranges) was also a shift to less-preferred foods, foods that, to me, tasted worse. For example, I preferred the taste of orange juice to the taste of oranges. Processing costs money, so the output (processed food) must be worth more than (must be preferred to) the input (the same food before processing). One way to increase preference is to increase the strength of flavor-calorie associations. Moreover, common forms of processing increase flavor (e.g., juicing, adding spices, adding fat) or cause physical or chemical changes that speed up digestion (e.g., mashing, shredding, heating). Increasing flavor increases CS intensity, a standard way to increase the strength of CS-US associations (Mackintosh 1974). Speeding up digestion reduces the CS-US interval, another standard way to increase associative strength (Mackintosh). So common forms of processing would be expected to strengthen flavor-calorie associations.

The new theory does *not* readily explain Example 6, that drinking lots of water seemed to lower the set point. However, the formation of a flavor-calorie association requires that the flavor be remembered until the calories are detected, which happens many minutes later. Drinking water may weaken the flavor memory. Studies of flavor-aversion learning have found that ingestion of food or drink between a flavor CS and a sickness-causing US reduced the effect of the CS-US pairing (e.g., Bernstein 1980; Fenner & Bernstein 1984).

The evolutionary argument and the theory's ability to easily explain the effect of less-processed food persuaded me to take it seriously. I wondered what it implied about how to lose weight. Too little is known about flavor-calorie learning to predict with confidence the effect of most dietary changes, such as eating fish instead of beef. However, the theory suggested where to look for ways of losing weight – namely, among dietary modifications that reduce flavor or slow digestion.

3.3.2. Birth of the idea. What dietary modifications slow digestion? The *glycemic index* of a carbohydrate-containing food indicates how quickly blood glucose increases after eating the food: the sooner the rise, the higher the glycemic index. Mashed potatoes and bread, for instance, have high glycemic indices (Miller et al. 1996). It is obviously a good measure of the speed of carbohydrate digestion. Low-glycemic-index foods surely generate slower calorie signals, and thus longer CS-US intervals, than do high-glycemic-index foods. This suggested that reducing the average glycemic index of my diet would lower my set point and therefore lower my weight.

Legumes, especially beans and lentils, have low glycemic indices. So the new theory predicted that eating more beans and lentils, in place of foods with higher glycemic indices, should cause weight loss. Hearing this prediction, a friend said her boyfriend had been much thinner in high school, when he ate a lot of beans and rice (Joyce Friedlander, personal communication, August 17, 1996).

3.3.3. Test. At the time, my main sources of protein were fish and rice. The next day, I started eating beans and rice instead of fish and rice. I also started eating lentils more often. Figure 27 shows the results. I lost weight quickly for a short time, about 3 kg in 20 days, not counting 7 days out of town when I went off the diet. I had no trouble following a low-glycemic-index diet and never regained the lost weight.

3.3.4. Related results. An interest in heart disease risk factors led Ludwig et al. (1999) to study about 3,000 adults for 10 years. During Year 7, the subjects filled out a diet questionnaire. Amount of dietary fiber correlated with weight at Year 10 (more fiber, less weight) as well as weight gain from Year 1 to Year 10 (more fiber, less weight gain). The difference in weight between white persons (I am white) in the

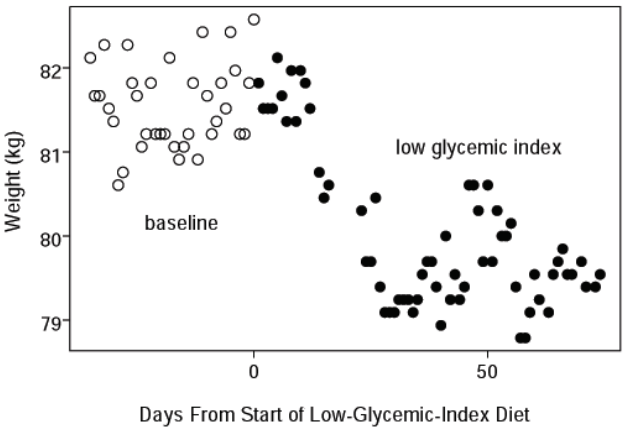

Figure 27. Effect of a low-glycemic-index diet on weight. The diet began July 18, 1996.

highest and lowest quintiles of fiber intake was 3.6 kg, close to the effect seen in Figure 27. Kromhout et al. (2001) found a negative correlation between fiber intake and BMI. Fiber content and glycemic index are negatively correlated (Wolever 1990); beans, for instance, have a low glycemic index and lots of fiber.

In an experiment, rats fed a high-glycemic-index diet for 32 weeks weighed more than rats fed a lower-glycemic-index diet (Brand-Miller et al. 2002). Roberts (2000) reviewed several studies that found that after a high-glycemic-index meal subjects ate more over the next 24 hours than after a lower-glycemic-index meal. Brand-Miller et al. (2002) reviewed other evidence suggesting a connection between glycemic index and obesity.

3.3.5. Discussion. My weight loss was not large but it was more than the weight loss produced by a low-fat diet in several studies (Willett 1998). It was impressive support for the theory, not because the theory provided the only explanation – it is unlikely that any weight change has only one possible explanation – but because the theory had led me to try a low-glycemic-index, and because, like the effect of processing, it differed from the evidence that suggested the theory.

By 2002 (when this was written), the idea that a low-glycemic-index diet will cause or is likely to cause weight loss had gained many advocates, including scientists (Brand-Miller et al. 2002; Ludwig 2000; Roberts 2000) and authors of popular weight-loss books (Montignac 1999; Sears 1995; Steward et al. 1998). But even recently there was enough disagreement for Rolls and Barnett (2000) to state:

> The glycemic index is a poor guide for . . . weight management. The main reason is that there's little correlation between an individual food's glycemic index and that of a whole meal. . . . Even if it were more reliable, the glycemic index wouldn't help you lose weight or keep it off. "There's zero evidence that the glycemic index affects weight," says [Stanford University professor of medicine Gerald] Reaven. (Rolls & Barnett 2000, pp. 61–62)

In their introduction to a supplement about the glycemic index (five articles: four pro, one con), Ludwig and Eckel (2002) noted "on one thing all authors agree: There is a need for long-term clinical trials of low-GI [glycemic index] . . . diets" (p. 264S).

3.4. Example 8: Thick pasta → Temporary weight loss

3.4.1. Background. Example 7 made me believe that if I reduced the glycemic indices of my food even further, I would lose more weight. But it wasn't obvious how to do so.

3.4.2. Birth of the idea. In 1997, while reading a list of glycemic indices (Miller et al. 1996), I noticed that thinner pastas had higher values. The glycemic index of capellini (45 with glucose = 100) was more than the glycemic index of fettuccine (32), and thin linguine (55) was more than thick linguine (46). This made sense. Thin pastas have more surface area/volume than do thick pastas and thus more contact area/volume for digestive enzymes. They should be digested more quickly. For the same reason, thin pasta cooks faster than thick pasta.

The two correlations – thickness and glycemic index, and thickness and cooking time – suggested that an especially

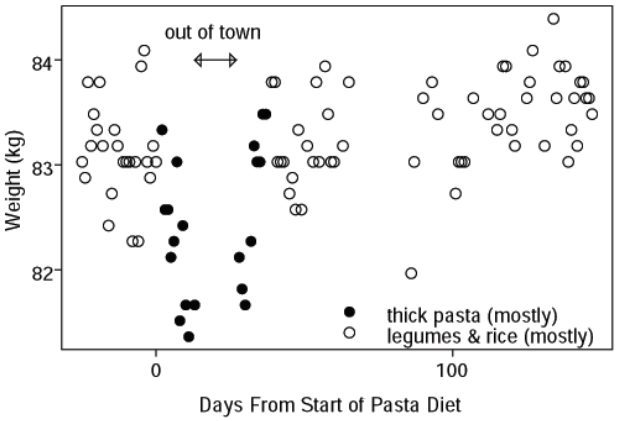

Figure 28. Effect of penne regate on weight. Arrows indicate days that I was out of town and not following the pasta diet. The pasta diet began September 26, 1997.

thick pasta, with a long cooking time, would have an unusually low glycemic index, perhaps lower than lentils (26–30), chickpeas (33), black beans (30), and brown rice (55), my main sources of carbohydrate at the time. If thick pasta had a lower glycemic index, Example 7 suggested that eating a lot of it might cause weight loss. The slowest-cooking pasta available at a local store, according to suggested cooking times on the packages, was penne regate (13 min), much slower than fettuccine (7 min) or capellini (2 min).

3.4.3. Test. I made penne regate the main source of calories at every meal. I ate it with tomato sauce, along with a lettuce salad. Figure 28 shows the results. I lost weight quickly at first, about 1.5 kg in 10 days. Then I went out of town for two weeks; while away, I ate an ordinary range of foods (no pasta). When I returned, my weight was about the same as when I left. I resumed the pasta diet. Contrary to what I wanted to happen, I started gaining weight and in about a week regained the weight I had lost. At this point I returned to my original low-glycemic, legumes-and-rice diet.

3.4.4. Related results. In other cases the eventual effect of a change in an ad libitum diet has been different from the initial effect. Willett (1998) described studies of low-fat diets where subjects regained some of the lost weight while still on the diet. Rat experiments have found that a dietary change that did not cause weight gain at first caused weight gain after about 1 week (Ramirez 1990a; Sclafani & Springer 1976).

3.4.5. Discussion. After I began losing weight I expected the weight loss to last a long time, as had earlier weight losses (cf. Fig. 1, lower panel, and Examples 6–7). That the weight loss was temporary suggests that expectations have little long-term effect on my weight.

The results are easy to explain. I had eaten almost no pasta for five years; I had stopped eating pasta when I reduced the amount of highly processed food in my diet (Fig. 1, lower panel). It is reasonable to assume that I had forgotten the flavor-calorie associations involving tomato-sauce-flavored pasta. In the beginning, the flavor of pasta (with sauce) was not associated with calories, so my set point went down when I ate it (it failed to raise the set point as much as the more familiar foods it replaced). After about a week, I had relearned the association, causing my set

point to increase when I ate it. One week is roughly the length of time used to teach flavor-calorie associations in experiments (Sclafani 1991).

It would be rash to say that no other weight-control theory can explain these results. But learning is at the core of the Pavlovian theory, unlike any other weight-control theory, and this sort of result – a novel food first causes weight loss, then weight gain, as strong flavor-calorie associations are learned – is a basic prediction of the theory. I did not go far enough to clearly establish the weight-gain part of the prediction.

3.5. Example 9: Sushi → Weight loss

3.5.1. Background. Bland food should cause weight loss, the new theory predicts, because weak flavors should produce weaker associations than strong flavors (Fig. 26). But the prospect of eating bland food was unappealing, so I did not test this prediction.

3.5.2. Birth of the idea. One afternoon in 1998, at an all-you-can-eat Asian restaurant, I ate a lot of nigiri sushi (perhaps 20–25 small pieces), along with ordinary amounts of vegetables (string beans, mushrooms), and fruit. I ate the sushi plain, without soy sauce or wasabi (horseradish). I ate nothing the rest of the day. The next day I was surprised to notice that I wasn't hungry at lunch time, my first meal of the day. I skipped lunch. I was astonished to notice that I was not hungry by dinner time. I ate a tiny dinner. The next day, I was still much less hungry than usual. Yet every day I was burning lots of calories. My exercise routine, which did not change, was walking uphill on a treadmill about 1 hour each day.

I realized that my weight-control theory could explain the lack of hunger. Sushi (without wasabi or soy sauce) has a weak flavor, so its taste should be only weakly associated with calories. I had been eating food with flavors stronger than sushi, so sushi might well raise the set point less per calorie. If a meal raises your set point less than usual, then at the next meal your set point will be less than usual, making you less hungry than usual. That was the short-term prediction: less hunger. The long-term prediction was that if I continued to get a large fraction of my calories from sushi, my hunger would remain diminished until my set point stabilized at a lower level, whereupon my appetite would return and my weight would stop decreasing.

3.5.3. Test. Sushi was appetizing, unlike my mental image of bland food. The sushi at the all-you-can-eat restaurants had about half the fish-to-rice ratio as the sushi from ordinary Japanese restaurants, but was still good. I continued to eat large amounts of sushi, along with vegetables and fruit. Before beginning this diet, I had been eating lots of legumes (with ordinary amounts of seasoning), rice, vegetables, and fruit (a very-low-fat diet), and no more water than needed to avoid thirst.

At the start of the sushi diet, I weighed 83 kg; over about 3 weeks, I lost 6 kg (Fig. 29). Although the diet was easy to eat, it was expensive, logistically difficult, and probably too narrow to be optimal. (And, I learned later, a high-fish diet can contain too much mercury; see Hightower & Moore 2003.) So I switched to a diet that emphasized Japanese-style soups (lightly flavored, lots of vegetables, small amounts of fish or meat) and brown rice. I hoped that this

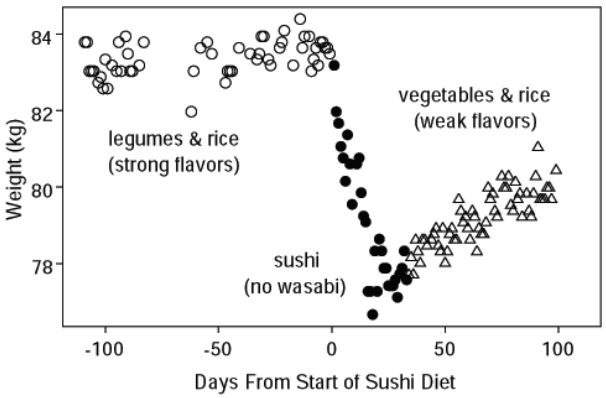

Figure 29. Effect of sushi on weight. The sushi diet began February 22, 1998.

would sustain the weight loss, but it did not. My weight slowly rose.

3.5.4. Related results. Bland food has caused substantial weight loss in several cases where subjects ate ad libitum. Kempner (1944) used a "rice diet" (p. 125) to treat the kidney disease and high blood pressure of two patients. One of them started at 69 kg (BMI 25) and lost 10 kg in 15 days; 8 months later his weight was even lower (58 kg). Another started at 74 kg (BMI 26) and lost 10 kg over less than 8 weeks. The diet was not entirely bland; in addition to rice, it included "sugar, fruit and fruit juices" (p. 125). Herbert (1962) had his food finely chopped and boiled three times to remove all folate. Starting at 77 kg, he lost 12 kg over 19 weeks. Cabanac and Rabe (1976)'s four subjects got all of their calories from Renutril (a Metrecal-like liquid food) for 3 weeks. Starting at 60–70 kg, they lost an average of 3 kg.

Unlike other cuisines, Japanese cooking does not use any strong flavors (Suzuki 1994), presumably to preserve the variety in taste provided by different fish. And the Japanese are much thinner than people of other rich countries (Intersalt Cooperative Research Group 1988). A migration study showed that this is due to environment, not genes (Curb & Marcus 1991). At the time of the Intersalt measurements, the difference between the average American adult male and the average Japanese adult male was 10–15 kg. In this experiment I lost 6 kg. I had already lost 5 kg from eating less-processed food (Fig. 1, lower panel) and 2 kg from eating low-glycemic-index food (Fig. 27), which suggests that I would have lost 13 kg eating sushi had I preceded the sushi diet with a typical American diet.

3.5.5. Discussion. Because it is unwise to eat lots of sushi, the discovery that sushi causes weight loss has only modest practical value, but it does add credence to the theory. The theory easily explains the effect, as described earlier, and it is quite different from other evidence for the theory. The effect of sushi is not easily explained by conventional weight loss ideas. Sushi is not high protein, low carbohydrate – a popular weight-loss prescription (e.g., Sears 1995). Nor is it a low-glycemic-index food because white rice does not have a low glycemic index. Sushi is low fat, but my previous diet was also low fat and, in any case, fat content makes little difference (Willett 1998). It might be argued that the weight loss was due to a reduction in the variety of my diet, which can have a powerful effect on weight (Raynor 2001). However, Example 8 (pasta), where I ate a very monotonous

diet, suggests that reduction in variety is not powerful compared to other factors.

The weight gain while eating Japanese-style soups was unexpected and disappointing. I had hoped to stay at the lower weight. Perhaps the weak flavors were not weak enough. Whatever the reason, the weight gain was more evidence, adding to Example 8, that expectations and hopes had little effect on my weight.

3.6. Example 10: Unflavored fructose water → Weight loss

3.6.1. Background. The lessons of Example 1 (eat food that is less processed) and Example 7 (eat low-glycemic-index food) were useful, but they put me at a BMI of 26 (slightly overweight). It was unclear how to lose more weight without replacing foods that tasted good with foods that generated weak flavor-calorie associations, such as raw vegetables.

3.6.2. Birth of the idea. In June 2000, I visited Paris. The food was excellent. I wanted to eat three meals per day but to my surprise and disappointment I had little appetite, even though I felt fine and was walking a lot. I realized that the new weight-control theory suggested an explanation: It had been hot and I had drunk two or three sucrose-sweetened soft drinks each day, about 630 kJ (150 kcal) each. All of them had been new to me because they were brands not available at home. The novelty meant that their flavors were not yet associated with calories and therefore would not have raised my set point. They had been sweet, of course, a familiar flavor that presumably *was* associated with calories. But maybe sweetness was effectively a weak flavor, I thought, and what I had observed was another instance, similar to Example 9 (sushi), of bland food reducing the set point.

3.6.3. Test. After the trip, I tested this idea. During a baseline period of 10 days, I ate normally. Then I began drinking fructose-sweetened water in addition to normal food. The Parisian soft drinks had probably been sweetened with a mixture of sucrose and fructose, but I used fructose alone because in rat experiments it produces weaker flavor-calorie associations than sucrose (Sclafani et al. 1993) and because it has a much lower glycemic index (23) than sucrose (65; Miller et al. 1996). I made the fructose water by mixing crystalline fructose and filtered tap water. The water was unflavored apart from the fructose so that no flavor-calorie associations would form aside from sweetness-calorie associations. I drank the fructose water in several portions over at least 2 hours (drinking a lot of fructose at once can cause diarrhea). To prevent the calorie signal generated by the fructose from becoming associated with the rest of my diet, I never drank the fructose water at meals or within an hour of meals.

The upper panel of Figure 30 shows my weight and fructose intake. On Day 1 of the fructose phase (July 20, 2000), I drank 360 ml (12 fl. oz.) of crystalline fructose, about 4,200 kJ (1,000 kcal), dissolved in 2 l of water. My appetite disappeared so completely that even one meal seemed like a lot. That was too much appetite reduction, so on Day 2 I drank no fructose water at all. On Day 3 I halved the daily dose of fructose to 180 ml (6 fluid ounces) of fructose, about 2,100 kJ (500 kcal), keeping the amount of water (2 l/day) the same. I continued to feel no hunger at all between meals

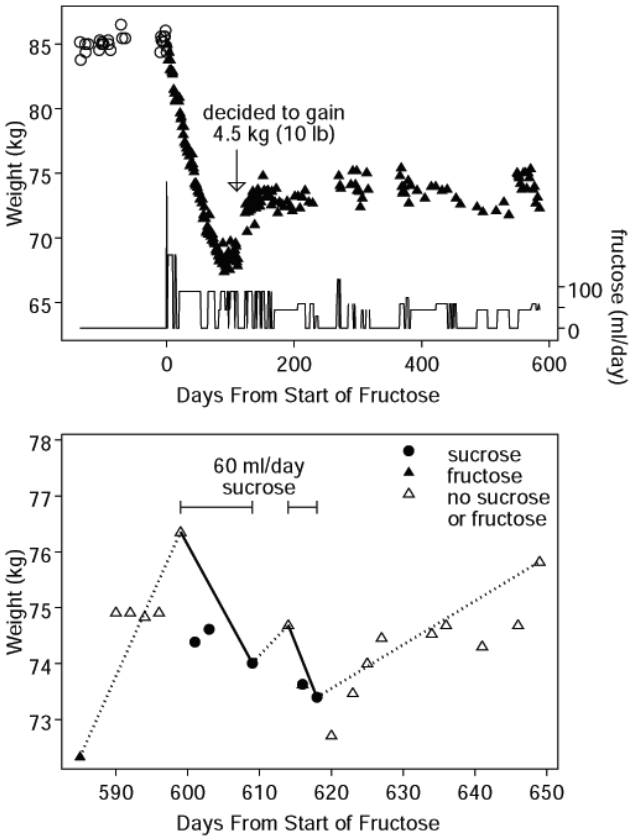

Figure 30. Effect of fructose water on weight. The fructose began July 20, 2000. Upper panel: Basic effect. Lower panel: Effect of sucrose. The treatment attached to each point (sucrose, fructose, and no sucrose or fructose) is the treatment during the days between that point and the previous point. The lower panel begins at the date that the upper panel stops. Open symbols indicate periods without any treatment (no fructose or anything else). The weights were measured in the morning before anything was eaten. The lines show the effect of the treatments.

even though I was eating much less than usual. I stopped drinking the sugar water for several days and then halved the daily dose again, to 90 ml fructose (3 fluid ounces), about 1,040 kJ (250 kcal), and reduced to 1 l/day the amount of water I mixed it in.

My hunger between meals was zero or close to it. Eating seemed optional but plainly it was unwise to eat nothing so I ate about one small meal per day (in addition to the sugar water). After I started eating I developed some appetite and it was easy to continue, but not hard to stop. I tried to get adequate amounts of protein, fat, and calcium, and I took a multivitamin pill every day. I lost weight so easily that it seemed I could end up at whatever weight I chose. I decided to lose weight until I reached 68 kg (150 lb), a loss of 16 kg (35 lb). I thought that would be enough to convince everyone that it worked.

I reached the 68-kg goal and started eating enough to maintain my weight. After a few weeks, however, I noticed that negative comments ("are you sick?") were more common than positive ones. My BMI was about 20.5; less than 20.7 has been considered underweight (Whitney & Rolfes 1999). I took advantage of a trip to New York to gain 5 kg (10 lb), raising my BMI to 21.9. A daily intake of 45 ml of fructose (about 520 kJ, 125 kcal) seemed to be enough to keep me there (Fig. 30).

At my final weight, about 12 kg (26 lb) below my initial weight, I ate much less to stay there than I had eaten to stay at my initial weight. Before losing weight, I had eaten about 11 MJ (2600 kcal) per day, two large meals. Afterward, I ate about 5 MJ (1200 kcal) per day, one normal-sized meal plus two pieces of fruit and the fructose. (My daily activity, which never changed, included 30 minutes of aerobic exercise plus about 2 hours of walking.) At the lower intake, I noticed two desires I had not noticed before: a desire for taste, which could be satisfied with calorie-free tea, and a desire for crunchy food, which could be satisfied with crackers, carrots, popcorn, and apples.

I had no difficulty staying at the lower weight. After 16 months, it appeared I would never regain the lost weight, so I began to test my explanation of the weight loss. According to the explanation described below, the crucial feature of the fructose water was that it supplied calories with no flavor except sweetness. If so, other sugars should have the same effect. An alternative explanation was that something specific to fructose was responsible (e.g., Heacock et al. 2002; Watford 2002). If so, other sugars might not work. I tried sucrose. Of course, many people believe that sucrose causes weight *gain* (e.g., Steward et al. 1998). I stopped consuming fructose for 14 days in order to gain weight, then drank water with sucrose for 10 days (60 ml/day – one-third more than the maintenance dose with fructose). I stopped the sucrose for 5 days, then resumed it for 4 days (60 ml/day). Finally, I stopped it again for 31 days. The lower panel of Figure 30 shows the results. During the three no-sucrose phases I gained weight (0.29, 0.14, and 0.08 kg/day). During both sucrose phases I felt the familiar lack of hunger and I lost weight (-0.23 and -0.32 kg/day). The effects of no sucrose and sucrose were reliably different, $t(3) = 5.15$, one-tailed $p = .007$. That sucrose and fructose had similar effects implies that the low glycemic index of fructose was irrelevant. So does a comparison with Example 7, where a much larger dose of low-glycemic-index foods had a much smaller effect.

3.6.4. Related results. Two ad libitum experiments with humans have measured the effect of a large increase in sugar consumption. Tordoff and Alleva (1990) gave subjects cola-flavored soda sweetened with high-fructose corn syrup. The subjects consumed about 2200 kJ (530 kcal) of sugar per day for 3 weeks. They gained 0.5 kg compared to subjects given no soda or artificially sweetened soda. Raben et al. (2002) measured the effect of a large sugar intake for 10 weeks. Their subjects' daily intake of sugar was about 3300 kJ (790 kcal), 70% from flavored drinks, the rest from solid foods such as yogurt and ice cream. They gained 2.6 kg compared to subjects who ate the same foods artificially sweetened.

Rat experiments have often found that making (unflavored) sucrose water available, in addition to regular chow, causes weight gain. Ramirez (1987) concluded that this was because the sugar increased water intake. Rat chow is dry and hard, and water may speed up its digestion (reducing the CS-US interval).

3.6.5. Discussion. There is no precedent for losing so much weight so easily (Fig. 30, upper panel). A reduction in energy intake usually increases hunger (e.g., Doucet et al. 2000), but I felt much less hunger than usual. Subjects in the famous Minnesota Semi-Starvation Study (Keys et al. 1950) lost about 16 kg over 24 weeks due to caloric restriction. I lost 16 kg more quickly, but started at a greater BMI. Near the end of their weight loss, the Minnesota subjects found the experience very unpleasant and debilitating. Mayer (1968), who recommended caloric restriction for weight loss, made the same observation obliquely: "The victim of obesity must deeply want to become and remain more lean" (p. 153). To lose 18 kg by running requires running about 24 km (15 miles) per day (Williams 1997). Intestine-shortening surgery can produce similar weight loss but is difficult and dangerous (Grady 2000).

My weight loss cannot have been due to expectations because it was unexpected. I originally expected unflavored sugar water to have roughly the same effect as bland food. In fact, it was far more powerful. After I stopped losing weight, sugar water provided about 10% of my calories (600 kJ/day, roughly the amount in one can of Coke©) yet my weight was 12 kg less than where it began. In contrast, Herbert (1962) got all his calories from bland food and lost 12 kg. In Example 9, I got about three-quarters of my calories from bland food (sushi) and lost 6 kg. The weight loss cannot be due to a desire to lose weight, or participation in an experiment, or tracking my weight, or any other feature shared with Examples 6–9 because in Examples 6–9 I lost much less weight or no weight at all.

The new weight-control theory can explain the results with one added assumption: *Sweetness does not raise the set point.* To distinguish sweetness from other flavors is less arbitrary than it might seem because sweetness is unusual in other ways. Most flavors we learn to like, but a liking for sweetness is innate (Ramirez 1990b). The usual view has been that sugars taste good to encourage us to eat carbohydrates, an energy source. But Ramirez pointed out problems with this idea, including the fact that sugars are not energy-dense (by weight, they contain about half the energy of fat). Sugars occur in all plants because they are the form in which energy is transported. A better evolutionary reason for sugar preference, Ramirez argued, is that it increases plant consumption. This is quite plausible, given the modern-day benefits of plant consumption (e.g., Campbell & Junshi 1994). Supporting Ramirez's argument, Capaldi and her colleagues found that rats' preference for sweetness decreases when their hunger increases (cf. Capaldi 1996). Lucas and Sclafani (1996) found that food deprivation caused rats to prefer a fatty flavor over a sweet taste. "Apparently there is something aversive or unpleasant about sweetness when food deprivation is high," wrote Capaldi (1996, p. 67), which explain why dessert comes at the end of meals and why few cuisines include sweet main courses. When food deprivation is high, finding energy becomes more important. If sugars taste good for reasons unrelated to energy-seeking, it makes sense that sweetness would not affect the energy-storage system.

One prediction of this explanation was briefly confirmed: Unflavored sucrose water has an effect similar to unflavored fructose water (Fig. 30, lower panel). The theory's most striking prediction, however, is that to cause weight loss the sugar water must be unflavored. The results of Tordoff and Alleva (1990) and Raben et al. (2002), who found that *flavored* sugar water causes weight gain, suggest that this prediction is true. So does common sense. If drinking Coke caused dramatic weight loss, surely this would be well known. More evidence that fructose in other forms does not cause weight loss comes from Anderson et al.

(1989), who measured the effect of adding a substantial amount of fructose (840–1000 kJ or 200–240 kcal/day) to the diets of diabetics via "fructose cookies, a lemonade-flavored fructose drink, or crystalline fructose to be used as sweetener" (p. 338). In the hospital, this had little effect on weight. Outside of the hospital, subjects gained about 2 kg (4 lb) over 23 weeks.

The observation that my steady-state calorie intake after weight loss was very low is not new. (Steady-state intake is the intake that produces neither weight gain nor weight loss.) Persons in the National Weight Control Registry, who had lost at least 14 kg (30 lb) and maintained the weight loss for at least 1 year, reported an average daily intake of about 6 MJ (1400 kcal), which is remarkably low considering their high level of physical activity (Wing & Hill 2001). These values are roughly what one would expect from laboratory studies of metabolism in which the resting metabolic rate of persons who lost 10% of their body weight decreased by 15% (Leibel et al. 1995). In the Minnesota Semi-Starvation Study, weight loss lowered body temperature (Keys et al. 1950). Why does weight loss reduce metabolic rate? Leibel et al. (1995) quite reasonably took the metabolic changes to be part of the regulatory system that keeps a person close to his or her set point weight. "These compensatory changes may account for the poor long-term efficacy of treatments for obesity," Leibel et al. (1995, p. 621) concluded. But my experience and the experiences of persons in the National Weight Control Registry call this view into question. In both cases, *sustained* weight loss coexisted with a profound metabolic slowdown (assuming that the Registry members ate well more than 6 MJ/day before their weight loss). An alternative explanation is that metabolic rate increases with weight to allow you to eat more. More food brings you more essential micronutrients. Under Stone-Age conditions, I argued earlier, a person's amount of body fat reflected the price of food: the lower the price, the more fat. As the price of food went down, the price of additional micronutrients went down as well, so it made sense to "buy" more of them.

Self-experimentation allowed me to (a) test an idea that most people would consider very unlikely (that sugar drinks can cause weight loss), (b) notice an unexpected change (much more loss of appetite than expected), and (c) adjust accordingly (by lowering sugar intake). I can't think of another way to have turned my Paris experience into useful science.

3.7. Discussion of Examples 6–10

The new theory derived more from Ramirez (1990a), a masterful piece of open-minded science, than from my simple self-experiments; but the self-experiments had a big effect. Examples 1 and 6 are why I learned of the Ramirez results – they increased my interest in weight control. Examples 7–9 made the new theory more plausible and suggested that the forces it describes are important in everyday life. Example 10 is impressive evidence for the theory. Because it predicts that foods with unfamiliar flavors will cause weight loss, it caused me, when faced with a profound loss of appetite in Paris, to search my memory for foods with unfamiliar flavors. What came to mind were the many canned sodas I had drunk. Subsequent research strongly implied that this was the right answer. No other weight control theory would have suggested it. "It is no success for Newtonian theory," wrote Lakotos (1978, p. 6), "that stones, when dropped, fall towards the earth, no matter how often this is repeated. What really count are [confirmations of] dramatic, unexpected, stunning predictions." Example 10 is not quite the confirmation of a stunning prediction but it is quite surprising. The theory helped me find it, and it is much easier to explain if the theory is true.

The theory suggests new explanations of well-known phenomena. Many rat experiments have found that an increase in dietary variety increases weight (Raynor 2001). The new explanation is that including two foods in one meal allows cross-conditioning: The flavor signal produced by one food can become associated with the calorie signal produced by the other. This will generally produce stronger flavor-calorie associations than when the two foods are eaten separately. Consider a pastrami sandwich. The pastrami generates a strong flavor signal but a weak calorie signal (assuming just a few slices of pastrami); the bread generates a weak flavor signal but a strong calorie signal. Eaten at the same time they produce a much stronger flavor-calorie association than if eaten a few hours apart. There seem to be no direct studies of the effect of dietary variety on human body weight but there are some curious facts. One is the claim of Diamond and Diamond (1985) – without any evidence – that eating foods separately, a few hours apart, promotes weight loss. Assuming that the taste signal is forgotten within a few hours, when foods are eaten a few hours apart there is no cross-conditioning. Another curious fact is that the French are much thinner than Americans (Shetty & James 1994). The French smoke more than Americans, true, but it may also matter that the French, unlike Americans, eat only one dish per course. For example, Americans generally eat meat and vegetable during the same course but the French eat them at different times, a practice that reduces cross-conditioning.

Explanations of the worldwide increase in obesity (Seidell 2000) have included less physical activity, more energy-dense food, and less fiber (Kromhout et al. 2001; World Health Organization 2000). The new theory suggests a broader explanation: the worldwide increase in wealth. Wealth brings choice. Judging from rat experiments, when we are given a choice between two otherwise similar foods, we prefer the one with the stronger flavor-calorie association – which is the more fattening one according to this theory. An increase in wealth allows the purchase of better-tasting foods, which will often be foods with stronger flavor-calorie associations. According to the evolutionary argument for the theory (introduction to Example 7), our body-fat regulatory system is designed, like any good storage system, to make us fatter when food becomes cheaper. The worldwide increase in obesity indicates it is working properly.

4. General discussion

4.1. Summary of results

Twelve years of self-experimentation suggested several new cause-effect relationships and a new theory. The most important cause-effect relationships were:

1. Seeing faces in the morning on television caused a bidirectional change (worse, better) in mood (Example 2). The change began a half-day after seeing the faces and lasted about a day. The effect was strongest when the faces

were life-size at a conversational distance. Evening faces reduced the effect the next day. Travel across time zones reduced the effect for a few weeks.

2. Drinking unflavored fructose water caused weight loss (Example 10). A small amount of fructose per day sustained a large weight loss for more than one year, with no sign of regain. More extensive dietary changes, which included eating less-processed foods (Example 1), drinking lots of water (Example 6), eating a low-glycemic-index diet (Example 7), eating thick pasta (Example 8), and eating lots of sushi (Example 9), caused less or no weight loss, ruling out placebo-like explanations.

3. Standing reduced early awakening and increased how rested I felt when I awoke (Example 3). To produce these effects required standing about 8 hours per day.

4. Standing and morning light eliminated colds for more than 5 years (Example 5).

5. Breakfast caused early awakening (Example 1).

6. Morning light decreased early awakening and increased how rested I felt when I awoke (Example 4).

The new theory (background of Example 7) is about how diet controls weight. It assumes that the body-fat set point increases when you taste a flavor associated with calories: the stronger the association, the larger the increase. At other times, the set point declines. Examples 1 and 6 helped suggest the theory and Examples 7–10, especially 10, made it more plausible.

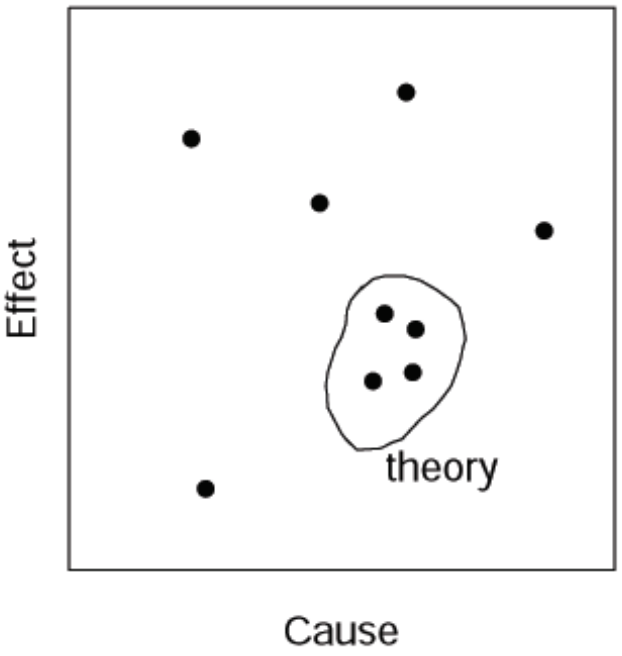

Figure 31. Cause-effect space. Each point represents a cause-effect linkage – a case where *x* causes *y*. A theory indicates a region of the space with an above-average density of points.

4.2. Value of self-experimentation

The number of new ideas, the strangeness of some of them, and their practical value all argue that self-experimentation is worthwhile. A new source of happiness (Example 2) can have far-reaching consequences; so can a new and powerful method of weight loss (Example 10). Both involved unusual treatments. Better health (Example 5) and better sleep (Examples 1, 3, and 4) are other popular and elusive goals that this work achieved in unexpected ways.

Self-experimentation helped produce plausible new ideas not only by generating ideas but also by testing and developing them. All three activities (generation, testing, and development) were important because in most cases the newly-conceived ideas, whether suggested by self-experimentation (Examples 1–5), self-observation (Examples 9 and 10), anecdotes involving other people (Example 6), or theory (Examples 7 and 8), were at first *not* plausible – at least, not plausible enough to be taken seriously by many others. Via testing and development (taking them further, adding detail) they became more plausible.

Use of self-experimentation to test ideas is fairly common (Altman 1987/1998); use of it to develop ideas is rarer. Ebbinghaus's (1885/1913) long series of experiments developed the idea that memory could be profitably studied that way. After animal experiments had shown that folate was a necessary nutrient, Victor Herbert (1962) ate a folate-free diet to determine the effects of folate deficiency on humans.

The novel feature of the present work was that self-experimentation repeatedly *generated* ideas. It did so by (1) producing "accidents" (unexpected observations) and (2) making me think. Accidents were behind most of the new cause-effect relationships. Thought-provoking results (the weight losses of Examples 1 and 6) helped me think of the weight-control theory.

Why did self-experimentation generate "accidental" discoveries at a high rate? Campbell (1960) proposed that creativity derives from a search process (trying many things, retaining just a few). Scientific discovery can be thought of the same way (e.g., Klahr & Simon 1999). To illustrate this, Figure 31 shows *cause-effect space*, with causes on the abscissa and effects on the ordinate. The space is mostly empty but here and there *x* causes *y,* marked by a point. Finding a new cause-effect relationship corresponds to finding a new point in the space.

Self-experimentation searched this space much more efficiently – that is, at a much lower cost/area searched – than conventional experimentation. The efficiency had several sources:

1. *It was easier to try new things.* Self-experimentation made it easier to test new treatments (new *x* values) and make new measurements (new *y* values). To discover and confirm a new cause-effect relationship generally requires novel experiments. In the present work many treatments were new, such as watching faces on television early in the morning (Example 2), standing many hours per day (Example 3), eating a great deal of sushi for weeks (Example 9), and drinking sugar water (Example 10). Doing such experiments with other subjects, even just one subject, would have been much harder.

2. *Self-monitoring made me sensitive to many changes.* Because we constantly monitor our behavior and feelings, we can detect many sorts of changes (sleep, mood, hunger, etc.) without effort. This allows self-experiments to detect changes on dimensions that are not the focus of interest. In contrast, conventional experiments can rarely detect change on a dimension not deliberately measured. In terms of Figure 31, this means that a self-experiment searches a *column* of cause-effect space, not just a point. The value of this feature was clear in Example 2 (mood), where some-

thing done to improve sleep in fact improved mood, with a surprising latency. Circadian-rhythm experiments sometimes measure mood (e.g., Boivin et al. 1997) but it is hard to imagine any conventional experiment detecting such an unlikely change. Examples 3, 4, and 5 also involved change on an unexpected dimension. Other researchers have had similar experiences. During a temperature-regulation experiment, Michel Cabanac sat in cool water for several minutes. Soon afterwards, when some hot water accidentally splashed on him, he was surprised to notice it felt pleasant. This observation began a long line of research about pleasure and homeostasis (Cabanac 1995).

3. *Long-term records (covering years) were kept.* Long-term records are hard to obtain without self-observation, and they help detect changes that might otherwise be missed. In Example 1 (breakfast), years of data were required to detect a change in sleep duration (Fig. 1). Likewise, long-term records made unexpected effects of standing (Examples 3 and 5) and morning light (Examples 4 and 5) easier to notice.

4. *Wisdom was cheaper.* Long-term research teaches what does and does not matter. Self-experimentation, because it is easier than conventional research, allowed such wisdom to be acquired more easily. Before Example 1, years of disappointing results had taught me that expectations had little effect on my early awakening. When I finally found ways to reduce early awakening (Examples 1 and 3), I knew that expectations were not responsible. In a similar way, Examples 6–9 acted as control experiments to show that the weight loss of Example 10 (sugar water) did not have a trivial explanation (e.g., more attention to what I ate).

Advantages 1 and 2 made it much easier to search cause-effect space, especially its unexplored regions. Advantages 3 and 4 made the detection process more sensitive (better able to distinguish signal from noise), so that fewer data were needed at each point.

Why was self-experimentation thought-provoking? Examples 1 and 6 put me in close contact with two cause-effect relationships that were not new to me. The results of Sclafani and Springer (1976) had made me believe that highly processed food is probably more fattening than less-processed food (Example 1). The notion that drinking water causes weight loss (Example 6) was one of many fringe ideas I'd heard about weight control. Learning that these cause-effect relationships held true for *me* was not a big surprise but surely made me think about them much more. They were hard to explain, so I was brought nose-to-nose with the inadequacy of current theory. Many have preached the value of direct experience. According to Halmos (1975) and others, "the best way to learn is to do" (p. 466), and the best way to learn a scientific finding may be to repeat or experience it. McGuire (1997), writing about idea generation, recommended "hands-on participation by the principal investigator in the research routine . . . rather than leaving these tasks wholly to assistants" (p. 26). An agronomist I know, while a graduate student, spent a great deal of time in his experimental plots. "My first couple of summers of experimentation, I would be either hoeing or crawling on my hands and knees through my fields, making measurements, for six to ten hours a day." His advisor believed that this would help him generate ideas. His field experience each summer has indeed been his best source of new ideas (Adam Davis, personal communication, December 23, 2001).

The low cost of self-experimentation allowed testing of highly implausible ("crazy") ideas. The notion that a diet high in water content would have wonderful effects (Example 1) was "crazy"; so was the idea that standing would cause weight loss (Example 3). Both were wrong, it turned out, but testing them paid off. I could never have tested them using conventional methods. The scientific literature, including textbooks, says little about implausible ideas, but they probably deserve more attention – after all, many "crazy" ideas (e.g., continental drift) have turned out to be true. Imagine all scientific ideas on a line, with each idea's position determined by the sum of everyone's belief in it. Ideas without a smidgen of belief are at one end; ideas that everyone believes completely are at the other end. The line has three regions: implausible, plausible, and true. Conventional research is almost always about ideas in the plausible region. Self-experimentation, this work suggests, is good for working with ideas in the implausible region. Long-term self-experimentation can find such ideas. Short-term self-experimentation can, via tests and development, push them to plausibility. At that point conventional research may take over.

A final virtue of self-experimentation, it might be argued, is that it showed the practical value of basic behavioral research. On the face of it, basic research has been of little help with the problems addressed here: happiness, depression, insomnia, the common cold, and obesity. Research about happiness, although quite interesting, has not yet led to experiments that increase happiness (e.g., Myers 1992). The common treatments for depression (antidepressants, ECT, psychotherapy) were developed by trial and error; understanding of the underlying mechanisms played almost no role (Valenstein 1998). Most instances of successful weight loss seem to owe nothing to research (Wing & Hill 2001), and similar statements could be made about insomnia and the common cold. However, the highly practical results reported here required, I believe, a great deal of conventional research, much of it from the past 30 years. When I was trying to understand why breakfast caused early awakening (Example 1), the results of Mistlberger et al. (1990) put me on the right track. Circadian-rhythm research (especially Wever 1979) and sociological research (Szalai 1972) led me to the strange observation that began Example 2; the many connections between depression and circadian rhythms (e.g., Wehr & Goodwin 1983) were one reason I took it seriously. The weight-loss method of Example 10 was found with the help of a theory that is so research-based (especially on Ramirez 1990a, and Pavlovian conditioning results) that I have yet to find a commonsense way to explain it. In terms of effort, the present work was a drop in the bucket compared to the previous work that made it possible. In this sense, self-experimentation was a catalyst that brought the prior work to fruition. The technology needed to do what I did became available long ago (with the exception of Example 2, which required cable TV and a VCR). But nobody did it, because the necessary ideas were missing.

4.3. Expectations

That expectations may influence results is an obvious worry with self-experimentation (Baer 1998; Rosenthal 1966). One way to address this concern is to equate expectations across conditions being compared (as in double-blind ex-

periments). This strategy has two subtypes: (1) *Zero expectations.* This applies to effects I had no reason to anticipate (i.e., surprises). Surprises can be fast or slow. The initial observation that began Example 2 (watching Leno and Letterman one morning made me feel very good the next day) took seconds to notice. The unusual loss of hunger produced by sushi (Example 9) took hours to notice. Within days I noticed that standing a lot seemed to reduce early awakening (Example 3). The loss of hunger produced by sugar water (Example 10) became more surprising as, over weeks, I reduced the amount of sugar and the absence of hunger persisted. The decrease in sleep duration shown in Figure 1 (upper panel) first came to my attention during data analysis many months later. (2) *Some expectations.* A subject in a double-blind experiment has some expectation of improvement, of course. The only work reported here that resembled a blind experiment was the portion of Example 3 (standing) that involved measurement of daily amounts of standing in tiny portions. At the end of a day I had little idea how long I had stood. Roberts and Neuringer (1998) describe another "blind" self-experiment, which compared caffeinated and decaffeinated coffee.

Another way to deal with the possible effect of expectations in self-experimentation is to measure the effect. When a treatment given some chance of working is tried and has no effect, this implies that expectations have no effect so long as it can be assumed that the effect of the treatment and the effect of expectations were both not negative. In other words, if A (treatment) + B (expectations) = 0, and neither A nor B is negative, then B = 0. If you test a variety of treatments for which your expectations are equal, the effect of the expectations can be no larger than the smallest observed effect, if it can be assumed that no treatment effects were negative. This reasoning helped interpret Examples 1 (breakfast), 3 (standing), and 4 (morning light), where the previous long history of failure to find a solution to the problem of early awakening implied that expectations had no detectable effect. This reasoning also helped interpret Example 10 (sugar water) because Examples 6–9 showed that expectations alone could not produce anything close to the weight loss observed in Example 10. Example 8 (pasta), where the weight loss was temporary, implied that expectations had no long-lasting effect on weight.

Equating expectations and measuring their effect are "internal" (as in *internal validity*) ways of dealing with the issue. An "external" method is to compare conclusions from this work (where expectations might have made a difference) to conclusions from other work where expectations could not have played any role. Throughout the examples, the related results sections showed that the conclusions reached here were supported by other studies not involving self-experimentation. Table 2 summarizes this supporting evidence. It is more reason to think that expectations had no significant effect here.

This view of expectations agrees with what is known about placebo effects. Kienle and Kiene (1997) discussed the many methodological problems with the supposed evidence for the "power" of placebos. When those problems were taken into account, there was no longer any convincing evidence. "Having analyzed a total of 800 articles on placebo, we have not found any reliable demonstrations of the existence of placebo effects" (Kienle & Kiene 1997, p. 1316). Hrobjartsson and Gotzsche (2001) managed to find 115 studies that included a no-treatment control group in addition to a placebo control group, so that the effect of placebos could be measured. Combining these studies, they "found little evidence in general that placebos had powerful clinical effects" (p. 1594), although they detected small effects in the treatment of pain (where the concept of placebo effects originated) and with continuous subjective measures (such as the mood measurements of Example 2 and the rested ratings of Examples 3 and 4). They found no effect on objective measures (e.g., weight) or on binary measures (e.g., early awakening, colds).

4.4. Generality

Every study is $n = 1$ in countless ways (Tukey 1969) – one classroom, one set of materials, one time of day – so uncertainty about generality is not unique to self-experimentation.

One argument for the generality of the present results is

Table 2. *Outside support*

Example	Related results from conventional research	Section*
1 (breakfast)	Anticipatory activity generated by food	2.2.4
2 (faces)	Symptoms of depression; cyclical nature of bipolar disorder; links between depression and circadian rhythms; links between mood and circadian rhythms; social entrainment of circadian rhythms; correlation between isolation and depression	2.3.4
3 (standing)	High rate of osteoporosis in the United States, composition of leg muscles	2.4.4
4 (light)	Effects of light on circadian rhythms	2.5.4
5 (health)	Sleep-immune connections; timing of flu epidemics	2.6.3
6 (water)	None	3.2.4
7–10 (Pavlovian weight-control theory)	Evidence for the set point theory of weight control; Ramirez (1990a); Pavlovian conditioning experiments, especially CS and US pre-exposure effects; experiments that show flavor-calorie associative learning	3.3.1
7 (glycemic index)	Fiber-intake survey; rat experiment	3.3.4
8 (pasta)	Ad libitum diets where the initial effect differs from the long-term effect	3.4.4
9 (sushi)	Bland diets that have caused weight loss	3.5.4
10 (fructose water)	None	3.6.4

*Refer to the text section for details and references.

basic biology. Examples 1–5 suggest that certain environmental events (morning faces, standing, morning light) resemble vitamins, in the sense that they are required for basic processes to work properly. Likewise, other events (breakfast, evening faces, east-west travel) resemble poisons, in the sense that they cause basic processes to malfunction. A great deal of biological research teaches that this sort of thing does not vary from person to person; the necessary dosage varies, but the basic effect does not. Different persons require different amounts of Vitamin A but everyone needs some of it. What poisons one person will poison another. Examples 7–10 suggest the existence of a certain weight-control mechanism. Again, a vast amount of biological research argues that different persons do not have different regulatory mechanisms.

Another indication of generality is that for each of the problems studied or alluded to (poor sleep, depression, colds, obesity) the prevalence of the problem roughly matches the prevalence of its suggested cause: both are common. For example, poor sleep is widespread; so are breakfast, jobs that require sitting most of the time, and little exposure to morning sunlight. Depression is widespread; so is little exposure to morning faces. If this research had implied that something common caused a *rare* problem, it would have meant that something about me or about my environment was unusual. Phenylketonuria is a case where something common (foods with phenylalanine) causes something rare (mental retardation), and, indeed, persons with phenylketonuria are unusual. They have two copies of a rare allele.

The related results given with each example and summarized in Table 2 – results that, of course, came from other people or from animals – are yet another reason to believe that conclusions from the present results will be widely true. The results of Example 1 (breakfast), for instance, can be seen as an instance of anticipatory activity. A wide range of vertebrates show anticipatory activity. If a result generalizes across species, surely it will generalize within a species (from one human to another). The generality of Example 2 is supported by circadian-rhythm research and psychiatric research. And so on.

4.5. Idea generation

The present work supports earlier notions about idea generation. Many scientists have stressed the importance of accidents, which led to Examples 1–5, 9, and 10 (e.g., Beveridge 1957; Medawar 1984; Roberts 1989; Shapiro 1986). Root-Bernstein (1989) listed 43 "strategies for discovering" (p. 407) and some of them describe quite closely what I did, including "action creates results" (p. 409; Example 1 derived from years of trudging on in spite of failure), "try many things" (p. 409), "do what makes your heart leap" (p. 410; I cared a great deal about getting rid of early awakening), and, above all, "ignorance is bliss" (p. 417; when I began this research, I knew little about sleep, mood, health, and weight control, at least compared to experts). Likewise, some of McGuire's (1997) 49 heuristics held true, including the value of evolutionary thinking and "participating actively in the research process" (p. 5).

The present work illustrates how one discovery leads to others, a kind of self-catalysis. Discovery of new cause-effect relationships helps one induce a good theory; a good theory helps one deduce where to find new cause-effect relationships. The region labeled "theory" in Figure 31 indicates a region of cause-effect space with an above-average density of points. Finding one of the points within a theory helps you find others. The breakfast effect (Example 1) reminded me of the idea that our brains and bodies were designed for Stone-Age conditions, which led to Examples 2–5. Likewise, the effects of processing (Example 1) and water (Example 6) led me to the Pavlovian weight-control theory, which led to Examples 7–10.

If we add self-experimentation to the (short) list of methods of idea generation (already containing exploratory data analysis and combinatorial chemistry), we can see more easily what they have in common.

Exploratory data analysis (Tukey 1977) emphasizes visual displays, which are an excellent source of new ideas. Oliver (1991) advised would-be discoverers to "go for the spatial pattern," to look at the spatial pattern of almost anything (p. 78). An example is maps of county cancer rates in the United States (Jansson 1985). The colorectal cancer map revealed that "the northeast area with high colorectal cancer rates is very sharply separated along the Ohio River from the low-rate area to its south and less sharply but quite clearly along the Mississippi River from the low-rate area to its west" (Jansson 1985, p. 341). During the last Ice Age, glaciers stopped at this boundary. Perhaps sand carried by the glaciers buried protective nutrients deep underground. Tests of various possibilities led to a focus on the effects of selenium. This was not a wholly new idea (a few animal experiments had already suggested that selenium might be protective), but the discovery of wholly independent evidence gave it a great boost. Another thought-provoking observation was that Seneca County, New York, had much lower cancer rates than neighboring counties. Again, geological information helped generate the hypothesis that diets high in potassium reduce cancer, an idea supported by other comparisons.

Combinatorial chemistry consists of techniques that allow quick synthesis of a large number of chemicals with similar structures. It is now possible "to prepare in less than a week more compounds than were made in the whole previous history of chemistry" (Furka 2001, p. iii). The results are screened for the property of interest. Combinatorial chemistry has led to "the extreme acceleration of the process of discovery of active entities," wrote Seneci (2000, p. xi).

Visual displays of data, combinatorial chemistry, and self-experimentation all allow *rapid search of new spaces.* Visual displays of data allow the viewer to quickly detect any of many possible patterns – in effect, a very fast search through possible summaries of the results. To numerically test for each pattern, one by one, would take far longer. Combinatorial chemistry allows a quick search of many novel compounds for a property of interest. As discussed above, self-experimentation allows a rapid search of cause-effect space, especially unexplored regions. *Rapid search,* in these instances, means orders of magnitude faster than other methods. It is easy to believe that this property is the main reason each of the three methods generates discoveries at a high rate, and that any future method with this property will also be a good way to make discoveries.

Klahr and Simon (1999) noted that what scientists do when making discoveries often resembles a search process, which supports a connection between rapid search and idea generation. Another bit of supporting evidence comes from

the history of science, at least my nonexpert knowledge of it. New technologies often lead to scientific discoveries, of course, but in most cases years pass between the development of the technology and the discovery. Roentgen's discovery of X-rays, for instance, required photographic plates but occurred long after they became common. I know of only two exceptions to the long-latency rule: (1) Galileo was the first to use the telescope for astronomy and quickly discovered the hills and valleys of the Moon and the moons of Jupiter. (2) Leeuwenhoek made much better microscopes than his contemporaries and quickly discovered bacteria, protozoa, and many other things. In both cases, the time between improvement and discovery was much less than a year. Telescopes and Leeuwenhoek's microscopes both allowed rapid search of new spaces.

When combinatorial chemistry began, in the 1960s and 1970s, "leading academic chemists [made] comments such as 'this isn't science,'" recalled Pavia (2001, p. 52). Given the paucity of methods for generating ideas (Table 1), it *was* different from the scientific methods they had been taught. We might assume that methods of idea generation are rare because idea generation is mysterious, but the opposite may be closer to the truth: Idea generation is mysterious because methods of idea generation have been rare. With three methods, idea generation does not seem so mysterious: The methods have an obvious common element.

4.6. Innovation

Scientific discovery is one type of innovation and, judging by the present work, it follows rules that also apply to other types. Innovation often results from mixing (e.g., Bartlett 1958, p. 58; Koestler 1964, p. 236; Simonton 1999, p. 123). In the present work, a simple example is that the weight-control work (Examples 6–10) brought ideas from animal learning to a different research area. The crucial mixtures may have more than two components. Jacobs (1984) argued that for cities to prosper and diversify their economic activity (generating new products and services) a number of economic conditions must coexist. The present work mixes anecdotes, self-experimentation, theory, and conventional empirical research, in the sense that it has been guided by all four. In some cases, the "mixing" that generates innovation is only a meeting, a coming together of things that do not intermingle. Biological diversity is especially high along boundaries between two different habitats (e.g., forest and meadow), a phenomenon called *the edge effect* (Harris 1988, p. 330). Self-experimentation is on the border between anecdote and multi-subject experimentation and has features of both: the range and flexibility of anecdote, the repetition and logical clarity of experiments.

Innovation has been curiously linked with coherence, all the way from lasers (coherent light) to political movements (coherent action), perhaps because coherence is hard to achieve but, when it *is* achieved, creates a new source of power. The work described here took the events of one person's life, recorded day by day, and by imposing some uniformity pulled them together to answer scientific questions.

Other factors were also important. Smith (1981) noted that in a history of metallurgy he had written, "many of the primary sources . . . were objects in art museums" (p. 191) because many important metallurgical techniques were first used by artists. He attributed this to "the artist's search for a continued diversity of materials" (p. 191). To an artist, in other words, a new material has value because it differs from familiar materials. Whether the material is better in a standard metallurgical way (stronger, more flexible) matters less to the artist than it would to an engineer, yet an artist's interest in new materials has led to the development of materials that engineers have found useful. Likewise, a historian of art could be impressed by the effects of new technologies first used nonartistically – ceramics benefited from hotter, cheaper ovens, for example. That artists have different goals than engineers has helped engineers innovate, and vice versa.

In a small way, the present work suggests the same thing – that innovation benefits from diverse goals. I did not begin self-experimentation because I wanted to make discoveries. The main point of this article – that self-experimentation promotes discovery – was a surprise. The examples described here were preceded by 10 years of self-experimentation that, judged as science, produced nothing of value – no interesting discoveries, no publications. Had I thought of it as science, I would have been discouraged. But I didn't; I thought of it as problem solving (how to sleep better), so the lack of a scientific payoff did not matter. Yet that "nonscientific" work was surely essential to the work described here. During that period I solved various practical and intellectual problems associated with self-experimentation. I learned what designs worked. I learned how to record, store, and study the data. What may scare others away from self-experimentation, the worry that expectations will distort the results, turned out to be not a big problem. I learned this when I saw that my expectations were often wrong. Diversity of goals increases innovation, it seems safe to say, because it causes development of new and useful skills and materials (essentially, new tools) whose application to a problem is not obvious when the development begins.

Business students learn that incubation (time in a protected environment) gives new things time to develop. Research universities provide incubation because the research they support need not be profitable. In my case, tenure provided incubation within incubation because it protected me from the usual demands (publications and grants) of a research university. Just as incubation, a kind of tolerance, promotes innovation, intolerance hinders it. Jacobs (2000) argued that caste systems and other forms of discrimination retard economic development because they prevent certain jobs from becoming the seeds of new businesses. In societies that oppress women, for example, "half of their populations, doing economically important kinds of work, such as cooking and food processing, cleaning and laundering, making garments, and concocting home remedies, are excluded from taking initiatives to develop all that work [e.g., open a laundry] – and nobody else does it, either" (Jacobs 2000, p. 33). Likewise, failure to do or publish self-experiments prevents them from acting as the seeds of new conventional research. As mentioned earlier, belief that something is bad makes it harder to learn what it is good for – including what it could become.

ACKNOWLEDGMENTS

Some of the data described here were previously published in Roberts and Neuringer (1998) and Roberts (2001). I thank Bill Browning, Israel Ramirez, and Robert Root-Bernstein for copies of their work, and I thank Michel Cabanac, Franz Halberg, Saul Sternberg, and Peter Totterdell for helpful comments.

APPENDIX: TECHNICAL DETAILS

A.1. Example 1 (Breakfast)

To measure my sleep, I used a custom-made device about the size of a book. At night, when I turned off the lights to fall asleep, I pushed a button (*start*). Five minutes later, the device beeped. If I was still awake, I pushed another button (*awake*) in response. The awake button was large, lit from within, and easy to push. If I pushed it within four minutes of the beep, indicating that I was still awake, the cycle continued: Five minutes after the first beep, the device beeped again. If I failed to push the awake button within four minutes of a beep, the beeping stopped. This measured my sleep latency in 5-minute units. If I awoke in the middle of the night and did not get out of bed (e.g., to urinate), I pushed the awake button. In the morning, when I got out of bed, I pushed a third button (*stop*). The device stored the times that the start and stop buttons were pushed, and counted the number of times I pushed the awake button before and after falling asleep (two separate counts). Similar measures of sleep have a long history (Blood et al. 1997).

A.2. Example 2 (Faces)

As an example of what I watched in the beginning, on July 20, 1995, starting at 9:43 a.m., I watched *Politically Incorrect* (13 minutes), *The Real World* (54 minutes), *Road Rules* (19 minutes), and *The Simpsons* (6 minutes); on August 16, 1995, starting at 9:24 a.m., I watched *Newsradio* (23 minutes), *Frazier* (23 minutes), and *Comedy Product* (standup comedy; 22 minutes).

After I determined that the crucial stimulus was a face looking at you, I mostly watched *Washington Journal* and *Booknotes.* Those provided a high density of such stimuli.

A.3. Example 4 (Morning light)

The light source I placed on my treadmill for the main experiment of this example was an ordinary fluorescent fixture with four standard-length (1.22 m) 40-W fluorescent lamps covered by a diffuser. The lamps were General Electric F40SP65, which emit light with color temperature 6500 degrees Kelvin, resembling skylight. I was exposed to the light while I walked on a treadmill watching television. The light fixture lay on the arms of the treadmill, shining upward. My eyes were about 0.6 m horizontal and 0.6 m vertical from the center of the diffuser. The television was directly in front of the treadmill at roughly eye level.

The treadmill was near a window, so the fluorescent light, to be effective, had to be substantially brighter than light from the window. Measured with a light meter (Gossen Luna-Pro sbc) at the position of my eyes pointed at the center of the diffuser, the intensity of the fluorescent light was 1600 lux. During the experiment, light from the window (measured by putting the light meter at the position of my eyes and pointing toward the center of the window) was about 40 lux at 7:00 a.m., and about 100 lux 2.3 hours later. On the days I had taken walks, the outdoor light intensity at 6:30 a. m., measured by pointing the light meter straight up, was about 1000 lux. Light from the television was 10–20 lux.

Open Peer Commentary

How observations on oneself can be scientific

David A. Booth

School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. **d.a.booth@bham.ac.uk**
http://www.bham.ac.uk

Abstract: The design and interpretation of self-experimentation need to be integrated with existing scientific knowledge. Otherwise observations on oneself cannot make a creative contribution to the advance of empirical understanding.

Seth Roberts is right to argue that experiments on oneself are unduly neglected in contemporary science. Unfortunately, however, Roberts has misapplied scientific method in the studies that he describes running on himself.

There are four types of flaw in his claims to have evidence for some effects of visual exposure, movement practices, or food selection on expressed mood, perceived sleep, symptoms of a cold, and body weight.

1. Roberts' self-observations are contaminated by effects of his knowledge of previous observations that he made of himself. There is little or no point in self-replication when the phenomena depend on perceptible stimulation or controllable action.

2. Roberts' manipulations are confounded by known influences that may provide explanations of his observations that conflict with his hypotheses.

3. Contrary to the argument by Roberts, the unexpectedness of an observation makes no contribution to the strength of the evidence. This is because, if flaws 1 and 2 were avoided by considering only a single observation, the surprise becomes logically indistinguishable from mere coincidence. Worse, a lifetime spent looking for surprises will collect an increasing number of spurious concurrences. For example, this is the basis of the very high proportion of supposed intolerances for foods that, on testing, prove to be misperceived (Booth et al. 1999; Knibb et al. 1999).

4. In the case of his most "weighty" conclusions, Roberts' theory has been refuted by extensive prior research.

It is with some grief that I see Roberts spoil his case for self-study, because I began my research in molecular neuroscience and my education in cognitive psychology with experiments on myself.

My initial brain/mind interest was the neurochemistry of psychosis. In that context, I once ate nothing but a large bar of chocolate for lunch and analysed its metabolic products, to show that the origin of a compound seen more often in hospital patients diagnosed with schizophrenia was less likely to be in their brains than in the boxes of chocolates which visitors gave to them more often than to the nurses who served as the control group in that project (Booth & Smith 1964). The printer's block for the key figure still sits on my office shelf, labelled "In Memoriam: experimenter as his own subject" – although the nausea that I suffered after eating a half-pound of chocolate was clearly not fatal! Note that my metabolism (or the hospital visitors' gifts) could not be affected by my perceptions or my actions, given that I kept the chocolate down.

Eight years previously, I worked by myself through a little book of experiments in psychology (Valentine 1953). My memory is of laboriously training out the Müller-Lyer illusion and replicating the primacy and recency effects in recall of lists of words. Later, however, I found that I had a capacity for direction of attention sufficient for observable dissociative effects as autosuggested movement. So I became aware that directed forgetting could modulate that curve of deficits in serial recall. No one these days should estimate the size of an effect without comparing performance between people who know and don't know the correct hypothesis.

Some of my professional discoveries about hunger provide a basis for sympathy with Roberts' thesis but also for criticism of his examples. In some instances, I can't tell if personal experience stimulated the hypothesis, my theorising triggered the self-observations, both, or neither.

For example, Booth et al. (1970) provided the first (group) evidence that protein is better than carbohydrate in a meal at keeping hunger at bay some hours afterwards. The finding has been replicated several times (most recently by Long et al. [2000]); indeed, the original pair of experiments was limited, like all single studies, and so needed to be extended by different designs (very differently by French et al. [1992]). The effect, moreover, may be the key to low-carbohydrate diets (like Atkins'): weight loss occurs only when energy intake is lowered, and reduction of hunger by the raised protein content may enable this self-restriction to be better sustained (Bravata et al. 2003). The autobiographical twist on this is that I had gained the impression that meals based on rice, even when I had eaten enough to feel very full, left me hungry again little more than an hour later. For a long while now, I have believed that I can prevent this (other) "Chinese restaurant syndrome" by including enough flesh food of some sort in the meal. Yet I can't tell if this effect is self-experimental evidence for the hunger-delaying effect of protein or autosuggestion from – or valid application of – my theory of late mobilisation and utilisation of assimilated amino acids through the alanine cycle.

On Roberts' ideas about sleep, he pleads that the basis of his anticipatory awakening is "surely not expectation." Yet it is likely that he was aware that he was not going to have his usual breakfast: I regularly wake early when I know I have something unusual to do when I get up. Quite apart from autosuggestion, Roberts fails to allow for some obvious mechanisms. For example, eating fruit in the evening could induce earlier waking because fruit fills the bladder more than drier foods do.

The most disastrous moves made by Roberts deploy the notion of a "set point" for body weight. This concept of a reference value is simply redundant when there are opposing negative feedback functions (Booth et al. 1976; Peck 1976). Furthermore, even for body temperature regulation, the hypothalamus has only countervailing networks for heat production and dispersal – no 37°C-setting neurones. The urgent scientific issues about obesity are the mechanisms by which a person can most easily lose more energy than they gain during and between meals (Blair et al. 1989; Booth 1998): "set" points that move (!) divert attention from the real scientific problems. Similarly mind-numbing is the unoperationalised notion cited by Roberts that flavour-calorie associations increase "palatability" (Booth 1990; Conner et al. 1988).

Basic mistakes undermine these designs and interpretations. One example will have to suffice here: sucrose is a compound of fructose and glucose and so is useless as a control for fructose. Indeed, a lot of fructose without glucose is poorly absorbed and the resulting upset could reduce hunger (Booth 1989, p. 249). Roberts can swap anecdotes with his readers for a very long time, but scientific understanding is not advanced until a literature-informed hypothesis is tested between or within groups in a fully controlled design shown to be double-blind.

To conclude, personal experience can be a good way to get new ideas. Deliberate manipulation of the environment and keeping an eye open for unusual consequences may accelerate the generation of hypotheses. Yet the only way that science progresses with new ideas is to test novel hypotheses against existing theories in a competent design. Individualised analysis of complex performance (Booth & Freeman 1993; Booth et al. 1983; 2003) is also grossly neglected but requires adequate design and aggregated data. Don't compare conclusions; find out about mechanisms.

Dionysians and Apollonians

Michel Cabanac

Department of Physiology, Faculty of Medicine, Laval University, Quebec, G1K 7P4, Canada. **michel.cabanac@phs.ulaval.ca**

Abstract: There are two sorts of scientists: Dionysians, who rely on intuition, and Apollonians, who are more systematic. Self-experimentation is a Dionysian approach that is likely to open new lines of research. Unfortunately, the Dionysian approach does not allow one to predict the results of experiments. That is one reason why self-experimentation is not popular among granting agencies.

The title of this commentary is borrowed from a letter to *Science* by the Nobel laureate Albert Szent-Györgyi (Szent-Györgyi 1972). In it, Szent-Györgyi recalled that the terms Apollonian and Dionysian, proposed earlier by J. R. Platt (in a personal communication), reflect extremes of two different attitudes of mind that can be found probably in all walks of life, but apply especially well to scientific research. Apollonians tend to develop established lines to perfection, whereas Dionysians rely on intuition and are more likely to open new, unexpected lines of research

Self-experimentation, as demonstrated by Roberts, is evidently Dionysian. Roberts shows how the results of each experiment led to a new hypothesis, then to another self-experiment, and so forth. I can modestly confirm that I have also experienced a similar concatenation of ideas. For my M.D. thesis, I trained in the area of temperature regulation in relation to central nervous system physiology under J. Chatonnet (researching exclusively with dogs, which was still possible back then); then I did a post-doctorate under P. Dejours, a respiratory physiologist. Dejours proposed that I should check whether what he had discovered and termed "accrochage" – that is, the peripheral nervous signal for hyperventilation at the onset of exercise – would also take place with shivering. It did, but that is not my point here. Both of us served as subjects in an experiment with acute cold exposure aimed at arousing shivering. During these sessions, I experienced what J. Barcroft and F. Verzár had discovered and described as "basking in the cold": a sudden relief and relaxation between the bursts of shivering that accompany intense cold discomfort (Barcroft & F. Verzár 1931). This observation largely determined my future life in research, because after returning from Dejours' lab I started to study what caused "basking in the cold." This led to another fruitful experience. One day I had subjected myself to a hot bath in order to study the influence of core temperature on thermal comfort and was in a hurry because another subject was supposed to arrive in the lab any minute. While I was cleaning the bathtub with a hose, I became aware that the very cold water from the hose felt pleasant on my hand. The ice-cold water, usually unpleasant, felt good! It felt good because I was still hyperthermic and sweating. This was the beginning of a still-active line of research: first, on sensory pleasure, then, on other pleasures like taste: then, on phenomena like earning money or enjoying poetry. I have described the process of self-experimentation leading to new lines of research, and the sequence of events that determine lines of research in my book *La Quête du Plaisir* (Cabanac 1995).

Thus, my experience confirms Seth Roberts' experience. I congratulate *BBS* for giving a voice to such an unusual, and unpopular, view. The advantage of self-experimentation has been demonstrated in the target article. However, a profound disadvantage lies hidden in this type of approach, as recalled by Szent-Györgyi himself. The Dionysian approach of self-experimentation indeed favors "accidental" discoveries (by well-prepared minds), but it is impossible to predict what will be found, and what the next question will be. Yet, nowadays research is a profession, managed by professionals, with strict criteria. Large amounts of money are usually needed for research and this often comes from the public; hence, taxpayers must be confident that their contributions are well spent. In turn, granting committees tend to demand detailed descriptions of the research projects, with a cascade of experiments and their expected results predicted in advance. This is not

how the Dionysian brain works. Since the Dionysian approach is often undertaken when the outcome of experiments is unpredictable, self-experimentation is rarely popular with granting agencies.

Linking self-experimentation to past and future science: Extended measures, individual subjects, and the power of graphical presentation

Sigrid S. Glenn

Department of Behavior Analysis, University of North Texas, Denton, TX 76203-0919. **sglenn@unt.edu**

Abstract: The case for the value of self-experimentation in advancing science is convincing. Important features of the method include (1) repeated measures of individual behavior, over extended time, to discover cause/effect relations, and (2) vivid graphical presentations. Large-scale research on Pavlovian conditioning and weight control is needed because verification could result in easy and inexpensive mitigation of a serious public health problem.

Several years ago, Irene Grote, a former student of mine and now a research scientist at the University of Kansas, expressed her conviction that increasing the frequency of self-experimentation among behavioral scientists would contribute significantly to the scientific understanding of behavior. I was not convinced and countered that self-experimentation was a fine tool for personal development and self-understanding, but I doubted its relevance to advancing science. Roberts' target article happily proves Dr. Grote right and leads the way in using a time-honored and powerful research method to bring experimental rigor to the task of integrating the principles of evolutionary theory and the principles of learning.

The first feature of Roberts' research that demands my attention is his use of repeated measures, over extended time, of the behavior of an individual organism to experimentally analyze the variables affecting behavior. Roberts' use of long baselines followed by systematic and repeated manipulations of independent variables is characteristic of the work of Pavlov, Skinner, and virtually all behavioral researchers whose work with nonhuman organisms led to our current understanding of behavioral processes operating during the lifetime of individual organisms.

Roberts' use of this experimental method in self-experimentation has added to its usefulness in two ways. First, 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. Most people simply will not participate for years in someone else's scientific experiment, so the kind of long-term experiment that is usually possible only with nonhumans is now made possible with humans doing self-experimentation. Second, Roberts has extended the use of this experimental method to explore characteristics of human behavior that may have their origin in the history of the species. In doing so, he brings the experimental method to bear on hypotheses heretofore examined only in terms of correlational research, and his amazing graphs powerfully reveal the effects (or noneffects) of his independent variables. As pointed out by Smith et al. (2002b), "graphs constitute a distinct means of inference in their own right" and graphs "are uniquely suited for the coordination of scientific activities (experimental, numerical, and theoretical)" (p. 758).

The second arresting feature of Roberts' article is the potential importance, both theoretical and practical, of his theory that the set point for weight maintenance can be (and often is) changed via Pavlovian conditioning. If the role of Pavlovian conditioning is supported in the case of weight control, the theoretical importance of this finding extends well beyond the realm of weight change. The theory offers an entrée into the vast and almost unexplored borderland between evolutionary biology and principles of learning, and it brings experimental rigor to the task of integrating the principles of evolutionary biology and the principles of learning.

On the practical side, Roberts' data clearly call for a large-scale investigation under tightly controlled conditions to test the Pavlovian conditioning theory of set point. Should the hypothesis be confirmed, an easily implemented, inexpensive treatment will become available for losing unwanted weight and a major public health problem may be abated. The knowledge that conditioning accounts for some of the problems said to derive from "Stone Age" bodies living in modern environments should allow humans to engineer modern environments to make those environments work *for* them rather than *against* them, thus mitigating any mismatch between Stone Age life and modern life.

A third feature of the target article that stands out for this commentator is a sometimes subtle but pervasive message that our understanding of scientific method is limited by the grip of traditional formulations. As a graduate student, I was puzzled by textbooks on experimental psychology that proclaimed the researcher begins by "stating a hypothesis." I wanted to know how the experimenter came by the hypothesis and I suspected that it usually derived from casual observation (unsystematic induction), or worse, from a "theory" that was itself based on casual observation. This seemed almost to guarantee a great deal of wasted effort. In the target article, Roberts shows that self-experimentation is on a short list of methods that have been shown to generate scientific ideas systematically. By a concatenation of data derived from such experimental analysis and correlational data consistent with theories already supported by converging lines of evidence (theories such as the theory of natural selection), Roberts shows us a way to begin filling the "missing methods" cell in his table of scientific activities. And the way he goes about doing it is entirely consistent with the way the great behavioral researchers have always generated the ideas that eventually led to new discoveries: by systematically investigating what appeared to be a cause/effect relation in the behavior of an individual organism – often discovered while collecting data on something else.

From methodology to data analysis: Prospects for the *n* = 1 intrasubject design

Joseph Glicksohn

Department of Criminology and The Leslie and Susan Gonda (Goldschmied) Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat Gan, 52100, Israel. **chanita@bgumail.bgu.ac.il**
http://www.biu.ac.il/soc/cr/cv/glicksohn.htm

Abstract: The target article is important not only for *black-box* studies, but also for those interested in tracing cognitive processing and/or subjective experience (via systematic self-observation). I provide two examples taken from my own research. I then proceed to discuss how best to analyze data from the $n = 1$ study, which has a factorial design.

Self-experimentation, or what some term *systematic self-observation,* has a venerable history in psychology, going back to the Würzburg school (Humphrey 1951) and even earlier to the work of Ebbinghaus. Roberts eschews a focus on subjective experience, discussing, as he does, cause-effect relationships in behaviour. Yet, I would argue that the importance of his target article is not only for *black-box* studies, but also for those interested in tracing cognitive processing and/or subjective experience. I provide two examples from my own research.

Schacter (1976, p. 475) presents the "entoptic explanation" of hypnagogic imagery, which states that entoptics serve as the raw data for hypnagogia. To test this, we manipulated awareness of both entoptics and hypnagogia in the same observers – my two coauthors, Friedland and Salach-Nachum, serving in time-locked single-subject designs (Glicksohn et al. 1990–91). Systematic self-

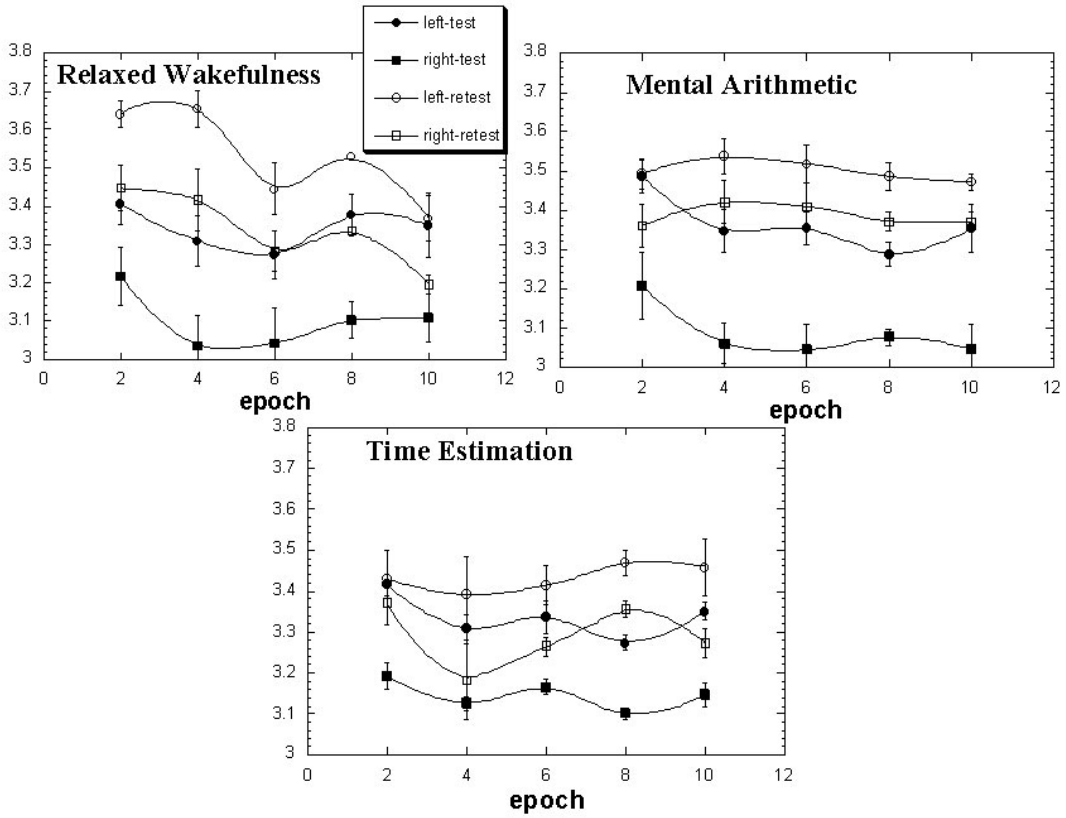

Figure 1 (Glicksohn). Mean alpha power values (after log transformation) together with SEM of one participant, as a function of epoch size, task, hemisphere, and session.

observation was conducted daily for 10 weeks, during which an experimental regime was followed: (1) exposure to a strong source of light for 10 minutes (week 2); (2) vigorous exercise (week 4); (3) cessation of smoking (week 6); exposure to drawings of entoptic phenomena (week 8). Vigorous exercise and cessation of smoking were incorporated because the observers had noted in the past that these factors tended to enhance the appearance of entoptics. The manipulations employed to explore hypnagogia were as follows: (1) use of an alarm clock to awaken approximately one hour after falling asleep (week 3); (2) exposure to monotonous sources of stimulation immediately prior to falling asleep (week 5); (3) exposure to unfamiliar surrealist pictures prior to falling asleep (week 7); exposure to monotonous sources of stimulation, and subsequent alarm-clock awakening (week 9). The entoptic hypothesis could not be supported: there was no correlation between the frequencies of incidence of the two. Furthermore, the content of the hypnagogic imagery was, in the main, not directly related to entoptics. In many respects, this study is similar to some of those described by Roberts. But note the difference in focus – not only on cause-effect relationships, but also on subjective experience.

My second example concerns the bipolarity of mood. One option is of reciprocal inhibition of polar opposites, whereby when one mood is dominant, its polar opposite is suppressed, but can later be released. To investigate the alteration of mood across time, two cooperative participants recorded daily mood (Glicksohn et al. 1995–96). Pleasant and unpleasant affect bore a lawful relationship, indicative of a cyclical, reciprocal inhibition of mood: assessed over the span of three months, they correlated negatively ($r = -0.64$ and -0.71, respectively). Furthermore, using spectral analysis, we found that the two were in opposite phase. Again, in many respects, this study is similar to some of those described by Roberts. But while Roberts reports on simple t tests to show reliable changes in sleep duration as a function of phase of diet, his data give clear evidence of cyclicity, better analyzed using spectral analysis.

This brings me to the issue of how best to analyze data from the $n = 1$ study, having a factorial design. There seems to be a confusion in the literature regarding the appropriate analysis of such data (see Crosbie 1995). Yet there is a readily available rationale for implementing an ANOVA with repeated measures, that is, to pool interactions in order to create an error term (Cox 1958, pp. 128–29). Let me demonstrate, using data from an unpublished EEG study.

EEG alpha is customarily quantified using spectral analysis. It is not clear, however, what should be the optimal length of sampling (epoch size). There is too wide a degree of latitude here, with different studies employing different epoch sizes, ranging from 1–4 seconds (e.g., Hoptman & Davidson 1998) to 15–30 seconds (e.g., Wackermann et al. 1993). I conducted a parametric study entailing three tasks (with eyes closed), with online EEG recording: (1) restful wakefulness; (2) mental arithmetic; (3) time estimation. Two participants provided complete test-retest EEG, exhibiting conspicuous, well-regulated alpha within each of two sessions. The EEG was subjected to spectral analysis, employing epochs varying in size from 2 to 10 seconds. Alpha power density was extracted, defined as total power within that band (8–13 Hz); this was subsequently log-transformed to normalize the distribution. I focus on the data of one of these participants.

An ANOVA with repeated measures on epoch size, task, hemisphere, and session, using the 12 profiles (Fig. 1), was run to assess the stability of alpha power as a function of epoch size. In this analysis, a common error term was created by pooling all interactions. The effect for epoch size was significant [F (4, 51) = 4.33, $MSE = 0.004$, $p < .005$]. In short, for this participant alpha power was highest for the 2-sec epoch ($M = 3.39$) and lowest for the 10-sec epoch ($M = 3.29$). In addition, the main effect for both hemisphere and session was also significant [F (1, 51) = 127.56 and

113.21, respectively; $p < .005$], but that of task was not [F (2, 51) = 1.37]. (In a similar vein, Roberts has employed a one-way ANOVA with repeated measures, to determine whether standing had an effect on sleep latency.)

A second data-analytic approach is to employ a within-subject multiple-regression model (Lorch & Myers 1990), as has been previously implemented for the individual assessment of structural brain asymmetries (Glicksohn & Myslobodsky 1993) and for studying profiles of line-bisection performance (Soroker et al. 1994). Both hemisphere and session were dummy-coded, and task was defined by means of two dummy variables, using the eyes-closed baseline condition as a suitable reference. The regression was significant [F (5, 54) = 47.76, MSE = 0.005, $p < .0001$, R2 = .82], with significant contributions for epoch size ($b = -0.10$, $p < .005$), hemisphere (b = −0.19, $p < .0001$) and session (b = −0.18, $p < .0001$), and no significant contributions of either dummy variable indexing task. These results essentially replicate the previous findings.

Roberts might well consider more extensively analyzing his n = 1 data base, using the methods described above.

Self-experimentation and self-management: Allies in combination therapies

Irene Grote

University of Kansas, LifeSpan Institute, Lawrence, KS 66044.
grote@ku.edu **http://www.people.ukans.edu/~grote/**

Abstract: Self-experimentation is a valuable companion to self-management in the benefit of pharmaco-cognitive-behavior combination therapies. However, data on individuals participating as active therapeutic agents are sparse. Smoking cessation therapy is an example. Roberts' self-experimentation suggests trying more diversity in research to generate new ideas. This may inform current approaches to the cessation of smoking.

Self-management skills – usually referred to as self-regulation or self-control – are valuable tools relevant to nearly any setting where independence from medical, therapeutic, or teaching agents is a desirable long-term goal. Self-management can involve patients as decision-makers in the long-term outcomes of their treatment (see Grote [1997] for combination therapies of pharmaco-cognitive-behavioral dimensions).

Self-management skills emphasize prevention and, importantly, relapse prevention. Treatments which involve individuals as self-managing decision-makers, and therefore, as therapeutic agents, may empirically turn out to work better than those which do not do so. Given these premises, modern training of medical professionals and therapists seems to mandate instructor skills for teaching self-management, and to suggest that instructors practice self-study. Anecdotes, data, and conceptualizations in the history of science indicate that methodological self-study can contribute important outcomes to medical and other sciences (e.g., Altman 1987/1998; Neuringer 1984).

Where there is self-management, self-experimentation typically is not far away. Select academic programs started teaching students long ago the techniques of studying their own conduct or self-perception using scientific method (cf. Burkett 1987; Hoch 1987; Jacobs 1990). (For example: Jacobs exemplifies studying the effects of smoking, with self-experimentation, Burkett exemplifies studying ways to quit the habit, with self-management, and Hoch exemplifies combining both, studying effects and implementing cessation management.)

Self-experimentation motivates individuals' curiosity to *understand* cause-effect relationships resulting from variables that they implement systematically, without necessarily wanting to change their lifestyle. Self-management teaches individuals to change some aspect of their lifestyle by systematically changing relevant variables, usually packages of variables, without necessarily wanting to understand or analyze which variable contributes to the outcome. A combination of both approaches, self-experimentation and self-management, motivates individuals to understand the systematic changes in conditions under which they may be able to manage realistic changes in their lifestyle. Thus, Jacobs (1990) is interested in the effects of smoking on the human heart. He does not tell us whether he is interested in self-managing smoking cessation as a result of the findings of his self-experimentation. Burkett (1987) demonstrates that she manages to stop smoking. She does not convey whether she wanted, or needed to understand which component of her package contributed to her success. Hoch (1987) finds that he can quit smoking – to change his lifestyle – and can gradually reduce his criterion, nicotine content, while he counts monies not spent on cigarettes. Carrying Hoch a step further with self-experimentation would have systematically reversed change in nicotine level and money counting. But in clinical practice, reversals for the sake of scientific curiosity can outweigh short-term cost versus long-term benefit.[1]

For the point of Roberts' self-investigations, in each of the three examples on smoking cessation offered in this commentary, some degree of control was in the hands of the individuals who studied themselves, as well as some degree of ingenuity. Therapies on smoking cessation may benefit from the three types of self-exploration to confer some degree of control to participating individuals, and to come up with new ideas for therapy.

Smoking appears to be one of the hardest habits to kick, particularly for individuals with a history of use of other addictive substances: Effective therapies are hard to find: Recidivism is typically considerably higher than maintenance of abstinence. Choosing the best combination remains controversial (Petry & Simcic 2002; Shoptaw et al. 2002). The present comment proposes that the controversy may persist: Therapies may want to try to include smokers themselves as therapeutic agents and as decision-makers in their combination of pharmacological, physiological, and behavioral components. During self-experimentation, interesting findings may surface; at a self-management, level, participants may experience control over their own quality of life changes. Traditionally, individuals receiving therapy are not engaged as therapeutic agents, perhaps on account of the reluctance to accept any data from self-studies as a valid method. Roberts (1998) mentions this reluctance, and in fact likens it to the reluctance many have in accepting smoking as a hazard, despite evidence to the contrary (e.g., Centers for Disease Control [CDC] 1997).

Roberts' effective use of self-experimentation, his conceptual conclusions combined with traditional experimentation, may teach a valuable lesson for smoking cessation research: The persisting controversy about effective components of combination therapies may incite seasoned grant researchers, or any practitioner with an inquisitive experimental inclination, interested in research-based therapy, to invite individuals or groups of individuals to collect self-observation data and to brainstorm their observations and data in individual or in group settings, suggesting self-management, and spurring self-experimentation. Surprising new ideas may spring from such scientific adventure! Visual analysis of basic data designs for self-experimentation and for self-management can easily be taught, as results from self-experimentation replicated across individuals suggest (Grote 2003); accessible visual analysis may enhance motivation or self-analysis.

Roberts' self-studies demonstrate that diversity in scientific method and conceptualizations may promote "selection" for unusual results, and for a richer fabric of results, as confirming or disconfirming traditional techniques by unusual scientific techniques. The history of science supports Roberts' contention that observing diverse behavior *with diverse methods* is better than observing behavior of a limited group and with limited methods. Metaphorically speaking, based on evolutionary conceptions the strongest (data-based and outcome-effective) conclusions inform each other by way of different approaches, and may spur improved approaches. The criteria, based on diversity of approaches, become more stringent – similar to those operating in technology (Grote 1999).

I enjoyed seeing the visual presentations of Roberts' findings, in addition to the statistical formulae. The latter are useful for some scientific audiences, but esoteric to most individuals whom we might want to interest in data-based self-experimentation – or self-management in therapeutic combination technologies. Visual approaches are closer to common sense (Grote 2003). The presentation of Roberts' visual analyses would benefit if the scales of graphs were equal in units or maxima, where relevant, or if it were signaled to the reader when they are not and why they need not be; conventional techniques are available to this effect (see, e.g., Figs. 21, 24, and 30).

A tabular overview of the 10 years of self-experimentation in future reports by Roberts might help readers to catch up with him over several pages of text and references: For instance, it would help if his10 years of self-experimentation were organized by concurrent or subsequent data collections, represented by different figures and sections, and by categorizing what worked or did not work, compared to what did or did not confirm conventional beliefs and traditional scientific outcomes. Roberts' Table 2 in the target article serves as a nice example of this.

ACKNOWLEDGMENTS

This commentary was supported by NICHD grant # 5 PO1 HD18955-18. I thank Dr. Kim Richter, University of Kansas Medical Center, for sharing her interest in smoking cessation therapies, and for sharing some of the literatures motivating her work. My computer consultant Yvonne Channel, who established my website for the SESCC (Self-Experimentation Self-Control Communications) deserves particular thanks. Completion of the website remains in progress.

NOTE

1. Coincidentally, I happened to be correcting the copyedited version of the present commentary while on a flight back from Washington D.C. where, across the aisle from me, sat an academic colleague who had just managed to lose 50 lbs of weight in a brief time span, in the course of a therapy that he considered more effective and successful than the typical recidivism rates we hear about from weight and smoking cessation therapies. Upon my querying my colleague about his role in the successful outcome, data collection process, and analysis, he stated that he was not interested in an analysis of the variables which led to his success and that of 30% of his cohorts (though he did participate in data collection, and he found the daily feedback from his balance an important incentive). Perhaps my colleague provides an example of the degree to which "self-control" is separate from "self-management" (Grote 1997), and therefore from self-experimentation in Roberts' sense.

Self-experimentation chronomics for health surveillance and science; also transdisciplinary civic duty?

Franz Halberg[a], Germaine Cornélissen[a], and Barbara Schack[b1]

[a]*Halberg Chronobiology Center, University of Minnesota, Minneapolis, MN 55455;* [b]*Institute of Medical Statistics, Computer Science and Documentation, Friedrich Schiller University, D-07740 Jena, Germany.*[1]
halbe001@umn.edu corne001@umn.edu
http://www.msi.umn.edu/~halberg/

Abstract: Self-surveillance and self-experimentation are of concern to everyone interested in finding out the factors that increase one's risk of stroke from <8% to nearly 100%; one also thereby contributes to transdisciplinary science.

Why chronomics? Whenever there are inter-individual differences in response, clinical trials on groups that do not consider such differences cannot solve what only the individual can do cost-effectively, such as finding out whether one's blood pressure (BP) responds to an increase in sodium intake with a rise, with no change, or with a decrease. Eventually, special institutions may be designed for the support of each individual's chronomics, that is, for the transdisciplinary monitoring of *chronomes* (time structures; from *chronos,* time, and *nomos,* rule) of biological variables, and for their interpretation, archivization, demographic analysis, and follow-up for outcomes. Recently, a particular individual's record sent to an international project on "The Biosphere and the Cosmos" (BIOCOS; corne001@umn.edu), coded to guard confidentiality, has become part of a promptly accessible database for both personal, individual needs and society's requirements. What was started by individuals and small groups as a self-experimentation in chronobiology, is currently available as a service by BIOCOS and could become a public system of planned surveillance and archivization of a person's "womb to tomb" chronomes (Halberg et al. 2001; 2003)[2]. Alterations of a standard deviation of heart rate (HR) and/or of a BP circadian rhythm's amplitude or acrophase can signal risk elevations, prompting countermeasures. Intervention prompted by risk elevation rather than by overt disease has been called prehabilitation.

Automatic ambulatory monitoring equipment is available at a 85% discount through the BIOCOS system; this project's Minnesota center analyzes, manages, and exploits all the available information, thus rewarding all participants and not just the individual providing the data. BP and HR are monitored automatically, ambulatorily, without vexing electrodes; or they can be manually self-measured around the clock and calendar, for long spans. Some opinion-leading physician-scientists suffering from hypertension, including a head of the clinical center at NIH, self-monitored their condition for decades from diagnosis to the fatal event. For those who have an over-the-threshold BP variability or an under-the-threshold HR variability, the risk of strokes and other cripplers rises from less than 8% to about 40%, even when the 24-hour average BP is acceptable. When both outlying variabilities coexist, the stroke risk is 80%. The risk is very high also with a further coexisting over-the-threshold difference between systolic and diastolic BP (pulse pressure), as shown in Figure 1. Excessive BP variability can be reduced and its excess eliminated, sometimes by rescheduling the timing of a drug without altering the dose (Halberg et al. 2003).

Prehabilitation. Public concern for clean air and water and clean, safe streets can be matched by striving for a safe circulation, to save the cost of rehabilitation by self-experimentation. Self-surveillance, initiated and implemented by an educated public, benefits health and science and is cost- and litigation-free. Preventive health maintenance – prehabilitation that is neither a commodity for sale nor a birthright – is an obligation to oneself and to society. A governmental cartographic institution for chronomics by prehabilitation may eventually reduce the need for agencies and resources instituted by law for rehabilitation.

Science. With goals in health care, having one agency serving both chronomics and the individuals it involves could make a transdisciplinary contribution to science; for instance, by making physiological recordings to match routine monitoring in atmospheric and solar-terrestrial physics, which has been systematically going on for centuries on Earth, and for decades via satellites. Newest is a wobbly average near ~1.3-year component in around-the-clock records for up to 35 years in the BP and HR of human "test pilots," nearly matching Richardson's also-wobbly ~1.3-year variation in the solar wind, recorded via satellites (Richardson et al. 1994). That a certain religious motivation shows an approximately 21-year cycle with latitude-dependent characteristics (Starbuck et al. 2002), with a period length found not only in homicide and war but also in Hale's bipolarity cycle of sunspots, should prompt self-experimentation in order to clarify the underlying mechanisms. Time-varying amplitude-unweighed phase synchronization, as a method developed by one of us (Schack), reveals the novel phenomenon of transient multifrequency phase synchronization between a biomedical series (daily incidence of mortality from myocardial infarction) and a host of aligned putatively influential physical variables. This new time-varying multifrequency phase synchronization, illustrated in Figure 2, is of basic interest

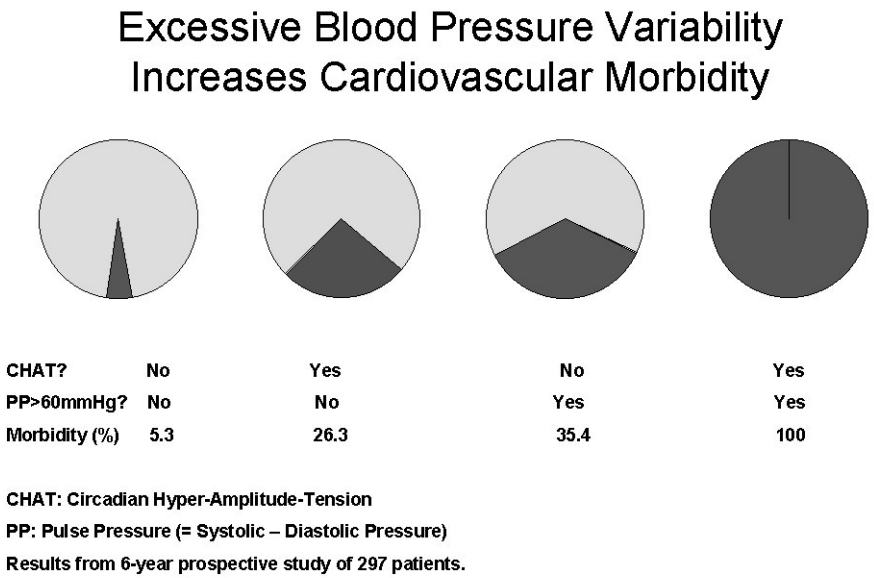

Figure 1 (Halberg et al.). Altered blood pressure (BP) dynamics (chronomics) raise the risk of cardiovascular morbidity from 5.3% to 100%.

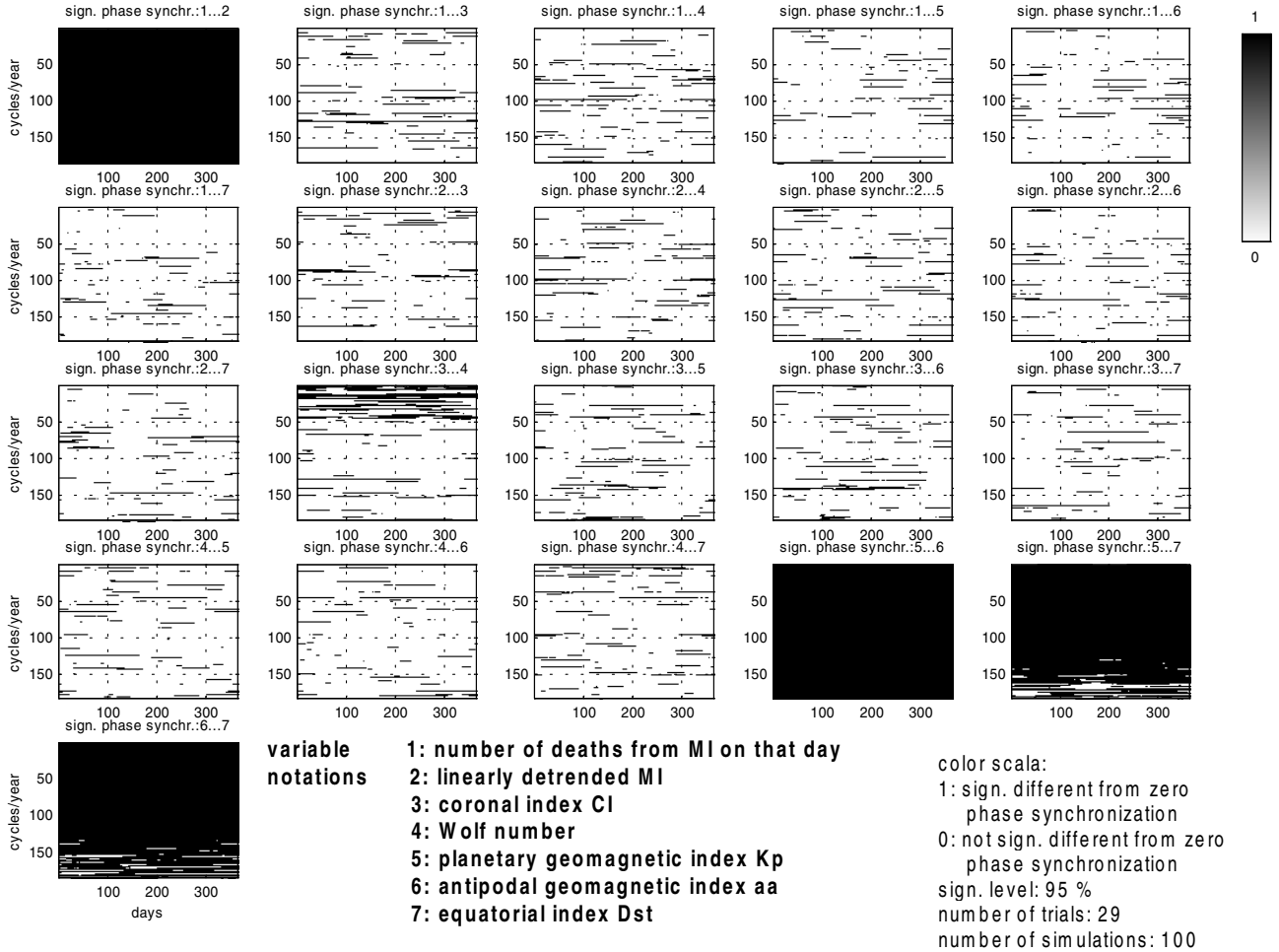

Figure 2 (Halberg et al.). Non-zero time-varying phase synchronization between the daily incidence of mortality from myocardial infarction in Minnesota during 1968–1996 (29 years) and different physical environmental variables, reflecting, among others, the influence of the sun and/or that of geomagnetism, determine any mediating (such as resonating) frequencies at any given time, and assess the extent of consistency of any frequency-dependent effect over time. Black rectangle on top left shows phase synchronization between the original and detrended data on myocardial infarctions, whereas the last three, mostly black rectangles show the also anticipated phase synchronization at most if not all tested frequencies among three geomagnetic indices, anticipated to be closely related.

in itself and may also provide useful applied information, for example, for shielding, replacing, and/or compensating magnetic fields as countermeasures. We had earlier learned that an approximate 7-day periodicity recorded in decades of around-the-clock measurements of HR is amplified in the presence and dampened in the absence of the same frequency in solar wind speed (Cornélissen et al. 1996). We owe these findings to the least squares of Gauss, who developed the method – to all of Europe's attention – to locate the "lost" asteroid Ceres in 1801. The method again proved useful to show that our heart and brain depend on our cosmos in more ways than sunlight and temperature (Halberg et al. 2003).

Conclusion. Self-surveillance and self-experimentation pertain to human health maintenance and the prevention of disease in its societal as well as classical aspects. A chronomic analysis of data from self-experimentation reveals interactions (time-specified feedsidewards) that account for the way we are affected by tangible physical environmental factors, so that one may try to elucidate mechanisms of diseases of individuals and of society for a broader-than-individual, also societal prehabilitation. Society has to "pay": in one way, by the family of affected individuals or by insurance premiums; in more ways than financially, for the care of every massive stroke that leaves the unfortunate survivors unable to care for themselves. Society faces an even greater bill, again more than financially, after each act of terror or war. This is the main challenge. The purpose of chronomics is to stray far beyond the scope of currently accepted limits for science and to tackle spirituality and "weather," not only on Earth but, since it affects us on Earth, also in extraterrestrial space. As to spirituality, "Things should be made as simple as possible but no simpler," as Einstein said. As to the weather, the effect of at least some unseen (nonphotic) aspects of the broader "weather" can be detected in ourselves. We gain some confidence from seeing wobbly signatures of geo- and/or heliomagnetics, as a near-week, a near-year and/or a transyear, in our physiology (BP and HR) and pathology (myocardial infarction).

NOTES

1. Dr. Barbara Schack died tragically on 24 July, 2003; memorial tribute in *Neuroendocrinology Letters* 2003, vol. 24, pp. 355–80. This note is also dedicated to her memory.

2. Each point made in this commentary is documented with figures based on data now published in Halberg et al. (2003).

Why does self-experimentation lead to creative ideas?

Todd I. Lubart[a] and Christophe Mouchiroud[a,b]

[a]*Université René Descartes – Paris 5, Laboratoire Cognition et Développement, CNRS UMR 8605, 92100 Boulogne-Billancourt, Cedex, France;* [b]*MGEN Foundation for Public Health, 75748 Paris Cedex 15, France.* **todd.lubart@psycho.univ-paris5.fr**
christophe.mouchiroud@psycho.univ-paris5.fr

Abstract: According to a multivariate approach on creativity, self-experimentation may well provide many of the conditions that allow for new ideas to occur. This research method is valuable in particular because the researcher's high level of participation in the search for a solution fosters the involvement of the necessary cognitive skills and conative traits.

Creative thinking involves the generation of new, original ideas that have value. Recent work on creativity proposes a multivariate approach in which cognitive, conative, and environmental factors combine interactively to yield creativity (Lubart 1999; Lubart et al. 2003; Sternberg & Lubart 1995). Cognitive factors include intellectual abilities such as selective encoding, selective comparison, and selective combination skills, as well as domain- and task-relevant knowledge. Conative factors refer to personality traits, such as risk taking, tolerance of ambiguity, perseverance, and openness, as well as intrinsic and extrinsic motivation. Environmental factors provide a setting that can stimulate and support the development of new ideas or, on the contrary, inhibit and devalue creative work.

Roberts proposes that self-experimentation is a powerful heuristic for creative thinking, and in particular, for generating new ideas. The main part of the target article presents the findings and new theories that resulted from self-experimentation. However, what makes self-experimentation special, compared to regular experimental methods? According to Roberts, being your own guinea pig has several advantages. First, it is cheaper and simpler to be your own subject. This can facilitate testing one's hypotheses and advance research quickly. Second, you can implement difficult-to-conduct methodologies, such as longitudinal research over months or years. Therefore, self-experimentation can facilitate creative thinking because it allows one to test difficult, perhaps crazy or long-shot hypotheses. Clearly, scientists tend to think twice about engaging resources in tests of hypotheses that offer little chance of success. The main benefit of self-experimentation is therefore to reduce the "opportunity cost" of testing creative ideas. This is not, however, a strong argument for the benefits of self-experimentation on idea generation; we can suppose that being able to test ideas cheaply allows one to entertain a wider set of ideas. But these ideas need to be generated at some point.

The potential contribution of self-experimentation to idea generation remains largely unexplored. We can, however, analyze Roberts' examples from the vantage point of research on creativity and build a case for the value of self-experimentation for creative idea generation. First, self-experimentation seems to facilitate selective encoding (noticing the relevance of information for one's task), selective comparison (observing similarities between different fields which clarify the problem), and selective combination (combining various elements of information which, joined together, will form a new idea). When a person has a specific question or issue in mind, he or she tends to see the world through "task-relevant" lenses.

The examples of self-experimentation proposed in the target article suggest that it is a 24-hour job to be both the subject and the experimenter. Consider Roberts' trip to Paris. Facing the culinary delights of French cuisine, he was curiously not very hungry. He noticed the oddity of this situation, perhaps, due in particular to his ongoing self-experimentation about weight loss. Then, he compared the culinary events of his Paris trip to those of his everyday life in the States. This selective comparison was facilitated, perhaps, because it was a within-subject affair. In fact, the novel taste of Parisian soft drinks was the distinctive element. Roberts happened to read prior to his trip a report about the novel taste of saccharin in experiments with rats. This previous knowledge made a selective combination process possible.

Self-experimentation can also facilitate creative idea generation by promoting certain conative factors for creativity. Most important, Roberts' examples of self-experimentation have a common feature: they all concern problems that Roberts considered to be personally important, namely, trouble sleeping, mood management, or losing weight. Roberts was therefore highly motivated to solve these self-engaging problems. He was willing to devote substantial time and effort to pursue possible solutions that could improve his life. He was intrinsically motivated, which has been shown to facilitate creative thinking. When he observed results in his personal quest that seemed odd, or contradictory, he lived with this state of ambiguity, and he continued to pursue his experiments because the possibility of finding a solution to his problems was worth the effort. For example, after testing several different breakfast menus, which seemed to have no effect on early awakening and were rather distasteful in some cases, he persevered because the elusive goal was worth pursuing.

Let's not forget that Roberts' self-experimentation did not occur in a vacuum. Friends, colleagues, and students provided anecdotes and suggestions that stimulated his self-experiments.

His knowledge of various work in anthropology, and animal research concerning dietary conditioning, contributed to the birth of his ideas and the methodology for testing them. His university position, with tenure, provided support and a certain degree of liberty.

In conclusion, it seems worthwhile to distinguish two kinds of self-experiments in Roberts' article. First, there are self-experiments in which a person tries different treatments, and varies more or less systematically the conditions to observe potential effects. Second, there are naturally occurring self-experiments, such as the change of beverages in Paris. These natural experiments happen all of the time. Most people do not notice the potential information offered by these "experiments," which may be a result of "chance" events. Being in the self-experimental frame of mind helps one to notice, to analyze these natural experiments, and to make good use of them. As Pasteur once said, "chance favors the prepared mind." Is it possible that self-experimentation serves creative thinking, in particular, because it leads a person to see the world in terms of a particular project, and to pursue this project actively or passively over a long period of time?

Self-experimentation as science

Harold L. Miller, Jr.

Department of Psychology, Brigham Young University, Provo, UT 84602.
harold_miller@byu.edu

Abstract: Examination of the target article for its relevance to the analysis of private behavior leads to three concerns: the absence of a new methodology for studying private behavior, the undisclosed possibility of interactions, and insufficient attention to the social context of idea generation. Regardless of these concerns, a larger issue remains: Can a science of $n = 1$ be credible?

Kudos to *BBS* for publishing a target article at once offbeat and provocative, and to its author for estimable perseverance. His decade-long efforts to solve the personal problems of early awakening and weight loss are ingratiating as is his canny ability to invoke references to the Amish, astronauts, sushi, and tenure, for example, without once mentioning serendipity (except in his references).

It is not readily apparent how the target article should be classified. Is it evolutionary psychology? Applied psychology? Learning? Motivation? Or does it belong to the philosophy of science? Roberts' report drew me initially by its potential to illuminate a longstanding lacuna in the experimental analysis of behavior, namely, the analysis of covert (also termed "private") behavior. B. F. Skinner's recognition of overt and covert categories of behavior (and the further categories of respondent and operant behavior) placed them in functional relationships with the categories of overt and covert environments. Skinner averred that experimental analysis would ultimately reveal causal links between covert behavior and the overt environment. In that linkage, covert behavior and overt behavior were likely to be correlated at best, and not form the causal relationship that is intuitively appealing. The issue of how to access private behavior for the purpose of measurement and thereby functional analysis was left open and remains underexplored.

Self-experimentation of the sort Roberts exemplifies and extols is tantamount to a science of $n = 1$. The researcher and his (in this case) subject are the same person. Psychologists will recognize the precedent established by Ebbinghaus; Roberts cites others as well. Like Ebbinghaus' interest in improving his memory, Roberts' interest was personally practical. Whether Roberts' two-cycle model of mood or his Pavlovian model of the set point for body weight will match the canonical status of Ebbinghaus' model of serial memory remains to be seen. What drew my interest was his assertion that the models were evidence of the particular efficacy of self-experimentation for the generation of ideas. Having ideas is private behavior.

My reading of the target article as a contribution to the experimental analysis of private behavior produced three concerns. The first is that Roberts produced no methodological breakthroughs in measuring the private events that were his desire for sleep, his restedness, his mood, his hunger, et cetera. He also utilized the A-B-A reversal designs familiar to experimental analysts, though never formally stating the stability criterion he invoked in moving from one condition to the next. Still, the extraordinary longitude of his data-gathering and its meticulous detail inspire awe. Together with clever procedures for presenting lights and faces, for recording wakefulness, and for instituting a lifestyle dominated by standing, not to mention a stoical submissiveness to flavor-impaired fare, they may even constitute heroic achievement.

The second concern pertains to Roberts' apparent neglect of the possibilities for interaction between the consecutive phases of his research. This is surprising, given his admirable alertness to the possible confounds posed by expectation and his care in dismissing them. Though his experimental designs began with a baseline condition, there was never any chance to return to a condition sans the cumulative effects of earlier conditions. For example, is it possible that the striking effects of fructose water on appetite suppression and weight loss were, at least indirectly, due to the effects of his earlier dietary experiments or, more remotely, to the fact that much of the waking day was spent on his feet and that he was sleeping soundly?

The inherently social context of the generation of ideas is my final concern. Several of the author's examples begin with parenthetical reference to a friend's comment or to some other incidental observation. An idea follows that soon takes on a life of its own and predicates the ensuing experimental series. The fact that the author lives alone and can configure his home life as he will, seems a distinct boon to the success of his experimentation. His considerable freedom to style his professional life in ways conducive to the research lends further advantage. In my opinion, he has given too little attention to these factors. Nor does he adequately acknowledge his reliance on friends and other associates who can be assumed to share his interests and on whom he is reliant for nascent ideas – and, I suspect, for ongoing support and encouragement of his work, if only casually. It is also important to remember his reliance on the published work of other scientists whom he readily credits.

These concerns are not meant to fault the author's exuberance or the promise of the kind of science he represents. Instead, they point up the possibility that self-experimentation is more complicated and less solitary than it might appear. It may also be more insular. The author's confidence that what he discovered on his own about himself will generalize to others has three bases: shared biology, commonality of incidence and cause, and interspecific convergence. Whether generality will be achieved remains to be seen. The economy of scale that, for the author, underlies the potent efficiency of self-experimentation may also constrain its generality. Consistent with my concerns, the weight of nuance may ultimately prevent sucrose water from capturing the market from Atkins and others.

Should failure of generality detract from self-experimentation? Can there be a science of $n = 1$? If the research is resolutely performed in the interest of hypotheses, and painstakingly adheres to the conventions of design, analysis, and peer-review, can it be acceptable if its findings invariably apply only to the behavior of the single subject who happens to live within the same skin as the experimenter? Does that suffice as science? Or does science require that specific findings always be replicable in others? If so, can the credibility of self-experimentation as science consist in the generality of the method – that others will self-experiment with the same admirable fervor and rigor as Roberts and with comparably salutary results? Whatever the verdict, the author must be commended for carrying on in the spirit that Skinner (1976) celebrated in *Walden Two,* where the character named Rogers ex-

claims, "I mean you've got to experiment, and *experiment with your own life*! Not just sit back – not just sit back in an ivory tower somewhere – as if your own life weren't all mixed up in it."

Can the process of experimentation lead to greater happiness?

Simon C. Moore[a] and Joselyn L. Sellen[b]

[a]*Department of Psychology, Warwick University, Coventry CV4 7AL, United Kingdom;* [b]*School of Psychology, Cardiff University, Cardiff CF10 3YG, United Kingdom.* **mooresc2@cardiff.ac.uk**
www.warwick.ac.uk/psych **sellenjl@cardiff.ac.uk**
www.cardiff.ac.uk/psych

Abstract: We argue that the self-experimentation espoused by Roberts as a means of generating new ideas, particularly in the area of mood, may be confounded by the experimental procedure eliciting those affective changes. We further suggest that ideas might be better generated through contact with a broad range of people, rather than in isolation.

Roberts claims to have found a novel association between television watching and his affective state at a later time. Despite Roberts' excellent experimental method, we would like to offer an alternative perspective that suggests the change in affect might be attributable to the process of experimentation itself.

Research into human affect has produced one seemingly robust and intuitive relationship: unexpected, rather than expected, increases in personal wealth elicit the greatest positive changes in one's affective state. In the case of money, people who receive an unexpected windfall report greater levels of happiness for up to one year after the event (Gardner & Oswald 2001), and unexpectedly finding a dime on a vending machine also elicits an improvement in participants' positive mood (cf. Schwarz & Strack 1999). To generalise, one could imagine that an unexpected increase in any valued commodity, be it money, status, even knowledge, could have the same effect. We argue that ideas represent high-value items for Roberts and that their discovery may lead to his greater happiness, rather than watching television in and of itself.

Roberts places a high value on ideas, as evidenced in his paper's introduction. Therefore, it appears reasonable to assume that, for Roberts, ideas may be considered analogous to other high-value items such as money and, as with those who receive an unexpected windfall, one might imagine that the discovery of a potentially fruitful association, or idea, could also elicit feelings of happiness. Although there has been no specific research on how the scientific process might elicit affective changes in scientists, it would appear intuitive to make such a connection, one that might offer an alternative explanation of Roberts' findings. Although our line of reasoning, that scientific discovery may elicit happiness, rests primarily on intuition and an analogous relationship between ideas and money, there is some indirect evidence supporting this claim. When the King of Syracuse instructed the mathematician Archimedes to investigate the material his crown was made of, it was not until Archimedes stepped into his bath and discovered that his bulk displaced an equal volume of water that he believed he had found a means of addressing the King's request. Not only was this an important scientific discovery, but it has also become synonymous with the happiness scientific discovery can bring. However, and importantly, it was Archimedes' belief that he had found a solution to the King's problem that elicited his happiness; as he leapt from his bath, he had not formally tested his theory. Could the prospect of a "Eureka" moment also have elicited happiness for Roberts?

Roberts became aware of an increase in his positive feelings *before* seeking its cause. The only plausible event that might be associated with his elevated mood appeared to be his television viewing on the preceding day (for the sake of argument, we will assume this was a random fluctuation in mood). The hypothesis to be tested became the relationship between television viewing and happiness: If watching television does increase happiness then, "Eureka," a new scientific discovery. This process, however, creates an important issue specific to the process of self-experimentation: Roberts, the participant, must have been aware of the hypotheses and aware of the manipulations he subjected himself to, and the value of such a finding. In other words, his anticipation of an important discovery may have led to an increase in positive affect, which was then falsely attributed to television watching (a similar argument can be applied to that period where positive affect was diminished, during the evening following watching television on day one, where there should be, according to Roberts' expectations, no evidence for a discovery).

We argue that the fluctuations in Roberts' mood may have been in consequence of the experimental process he engaged in, which, to generalise, may lead to questions surrounding the place of self-experimentation more generally. Being both observer and participant, we suggest, led Roberts the scientist to infer that feelings elicited by engaging in the scientific process were attributable to Roberts the participant, a claim that may go beyond research in mood. Are there means of generating ideas that might be less open to confound? We now suggest that there is already an abundance of ideas and that seeking means of developing new ideas is unnecessary.

One option to generate ideas may be simply to talk to people (Simon & Kaplan 1989). Although some important work has been conducted in solitary (e.g., Descartes 1637/1931), one does not have to go too far before finding people in applied settings who have an abundance of ideas, generated through observing real world behaviour, that are eminently researchable, but who have neither the time nor the resources themselves to test these ideas. One area proving fruitful is research that uses a form of protocol analysis (Simon & Kaplan 1989) with criminal offenders (McMurran & Sellen, in preparation). For example, there has been no systematic study into the motivations behind habitual offenders' decision to stop offending and their motivation to lead crime-free lives. McMurran and Sellen have started to examine, through interviews with offenders themselves and with practitioners in the forensic setting, the reasons behind offenders' switching their behaviour. Although the qualitative data is wide-ranging and broad, it is already providing novel insights (ideas) as yet not addressed in the experimental literature and that will ultimately lead to future experimental work. An advantage of this approach over self-experimentation is that there is little involvement of the researchers in recognising and describing areas of inquiry, but it is still close to the real-world behaviours to be researched.

In sum, we argue that self-experimentation, in the area of mood, may be confounded by the experimental procedure. We further suggest that ideas might be better generated through contact with a broad range of people in an applied setting where there is a great need for research and an already established means of analysis.

Experimentation or observation? Of the self alone or the natural world?

Emanuel A. Schegloff

Department of Sociology, University of California at Los Angeles, Los Angeles, CA 90095-1551. **scheglof@soc.ucla.edu**
http://www.sscnet.ucla.edu/soc/faculty/schegloff/

Abstract: One important lesson of Roberts' target article may be potentially obscured for some by the title's reference to "self-experimentation." At the core of this work, the key investigative resource is sustained and systematic observation, not experimentation, and it is deployed in a fashion not necessarily restricted to *self*-examination. There is an important reminder here of a strategically important, but neglected, relationship between observation and experiment.

Roberts' target article makes for compelling reading. Aside from its intriguing substantive results, it is a striking story of dedicated inquiry. Unhappily, the number of readers who are prepared to commit themselves to such a path in the future is surely limited, to say the least. So, admiring recruits aside, what other benefits and lessons are to be derived from this article? To my mind, one important potential lesson is obscured by the title. Rather than self-*experimentation* being the hero of this tale, self-*observation* is, and in a fashion which extends to include much naturalistic observation without limitation to "self."

As I read it, the text of the target article confirms the problematic observation with which it begins – that experimentation is best suited for *testing* new ideas, not for *getting* them. Most of the "experimenting" reported in the article is employed to test an observation or observed relationship, to chart its limits and variations, and so forth. But in most cases, a new direction is triggered not by the "experimentation" in the experiment, but by an *observation* (a "noticing" [sect. 2.5.2, para. 2], a "realization" [sect. 2.6.2, para. 1], a "noticing for the first time" [sect. 2.6.2, para. 3]), and, importantly, on a matter *other than what the experiment was oriented to examining.* Such telling observations which led to a reorientation of inquiry concerned not just the "value" of some variable or the strength of some relationship, but also the sort of variable or relationship that turned out to matter in the first place – as, for example, during experimentation with the effect of standing on weight loss (which was *not* exciting), Roberts' noticing that standing had an effect on sleep duration (target article, sects. 2.4.2, para. 3; 2.4.3, paras. 1–3, 11).

Roberts' account of the efficacy of self-experimentation in generating new ideas is that it produced "'accidents' (unexpected observations)" and made him think. In what sense were they "accidents," if they turned out to be naturally orderly phenomena? There are two senses:

1. Whereas "conventional experiments can rarely detect change on a dimension not deliberately measured" (sect. 4.2, para. 8), Roberts was able "to detect changes on dimensions that are not the focus of interest" (ibid.). Not having been measured, then, supplies one sense of "accident."

2. The very act of being attentive to one's surroundings and activities in a non-dismissive, open way allows anything potentially to "count." Thus: "Because I was recording sleep and breakfast on the same piece of paper, the breakfast/early awakening correlation was easy to notice" (sect. 2.2.2, para. 4). Neither of these sources of "accidents" has fundamentally to do with self-experimentation, though that is how Roberts happened to encounter them.

Indeed, these are not accidents at all, they are surprises; their "extraordinariness" is not a feature of their occurrence but of their being encountered – and *registered* – and *taken seriously* in a scientific sense – by the investigator. They occur because of a sort of inquiry in which what one thought before does not limit what one is allowed to "see and count" now. What Roberts was practicing was a form of orderly, disciplined, careful, and thought-through naturalistic observation, in which the very fact of close, careful observation allowed the connections and orderliness of everyday activities to become "remarkable." In his case, it was self-observation, but that does not strike me as criterial. There is quite a lot in human behavior that lends itself to this way of proceeding; unhappily it is only rarely taken seriously in contemporary psychology and cognitive science. In the spirit of Roberts' inquiry, I offer one episode from my own experience, with a suggestion for further reading.

Several years ago, a "friendly" psychologist/cognitive scientist refereeing a conference presentation of mine for a volume reporting the conference proceedings contrasted my "descriptive" and "post hoc" account with what more rigorous colleagues in cognitive science would want to see before having any confidence in it, but it seemed to me that the formal experimental testing that he proposed was insufficiently grounded in the target data, relied on the assessments of naïve (i.e., scientifically untrained) judges deploying the very "subjective" judgments for which trained, repeated, and systematic observation had just been called to task. (The friendly referee's comments and my responses to them appear in a postscript/appendix to my paper [Schegloff 1996], available at my website.)

The lesson to be learned from Roberts' work is institutional and disciplinary. If disciplines which are largely experimental in method granted those which are largely observational the courtesy of serious attention and uptake, many such "accidents" might fall into our collective laps. Once there, experiments could be used to test them. Of course, *non*experimental methods – including observational ones – can also be used to *test* new ideas, not just *get* them. But that is another commentary.

Ideas galore: Examining the moods of a modern caveman

Peter Totterdell

Institute of Work Psychology, University of Sheffield, Sheffield, S10 2TN, England. **p.totterdell@sheffield.ac.uk**
http://www.sheffield.ac.uk/~iwp

Abstract: A self-experiment by Roberts found that watching faces on early morning television triggered a delayed rhythm in mood. This surprising result is compared with previous research on circadian rhythms in mood. I argue that Roberts' dual oscillator model and theory of Stone-Age living may not provide the explanation. I also discuss the implications of self-experiments for scientific practices.

One means of achieving good sleep, mood, health, and weight is to live life as you might have done in the Stone-Age. This is the key claim behind Roberts' extraordinary set of self-experiments described in the target article. His examples of modern Stone-Age living include watching breakfast television, absorbing early morning daylight, delaying breakfast, eating sushi, drinking unflavored sugar water, and standing most of the day. I will comment on Roberts' more general point concerning methods for idea generation. First, however, I will focus on the explanation for what is perhaps the least intuitive finding from Roberts' set of experiments, namely, that watching faces on early morning television triggered a rhythm in his mood that commenced about 12 hours after stimulus and lasted approximately 24 hours.

An interesting feature of Roberts' delayed mood rhythm is that it was triggered by exposure to faces in the morning but was wiped out by exposure to faces in the evening. This suggests that the rhythm would be masked under normal conditions, because the timing of exposure to faces would normally be unrestricted. Research indicates that circadian rhythms in happy mood are also masked under normal conditions, but that they are revealed in specific circumstances, such as during depression (Haug & Wirz-Justice 1993), early infancy (Totterdell 2001), and extended sleep-wake cycles (Boivin et al. 1997). Two factors that appear to be important in revealing the happy rhythm are reduced reactivity to external events and the misalignment of the circadian pacemaker with the sleep-wake cycle. Roberts' abstinence from evening interaction and experience of sleep difficulties could therefore be relevant to his mood rhythm.

Closer inspection of Roberts' mood data suggests that both morning and evening faces caused a trough in mood about 18 hours after the stimulus. It is therefore plausible that the mood oscillation was dependent only on the social zeitgeber rather than being gated by a light-sensitive clock, as Roberts suggests. Roberts also uses his mood rhythm to propose that sleep and wakefulness are controlled by the joint action of a light-sensitive oscillator and a face-sensitive mood oscillator. Other research would suggest a different model. It is known, for example, that behaviors regulated by a circadian clock can feed back on the pacemaker (Wehr 1990b). Roberts' mood rhythm followed the same time course as the endogenous circadian rhythm in happy mood described by Boivin et al. (1997), so perhaps face exposure amplified that

rhythm. Models based on the joint action of a circadian clock and a sleep need process have also been very successful in predicting levels of alert mood at different times of day (Folkard et al. 1999). Mood researchers have usually differentiated between pleasure, arousal, and calmness in models concerning the structure of affect (Remington et al. 2000), and there is some evidence that circadian rhythms in happy and alert mood differ in form. Roberts' combined measure of happy, eager, and serene mood may therefore have confounded separate rhythms.

Roberts uses an evolutionary theory to explain his mood rhythm effect. Specifically, he proposes that seeing faces in the early morning resembles typical social interaction in the Stone-Age and that a mood rhythm triggered by such interaction facilitated Stone-Age communal living by synchronising people's sleep and mood. However, this post hoc teleological explanation is not entirely persuasive, and it raises difficult questions. For example, was seeing faces during Stone-Age sexual activity also confined to the early morning? Would it really have been advantageous for everyone in a community to sleep at the same time? Is a rhythm in irritability necessary to explain why people are usually irritable when woken? Why is a mood rhythm required to synchronise mood when there are many other more immediate unconscious and conscious processes known to bring about mood synchrony in groups (Kelly & Barsade 2001)?

Nevertheless, the delayed mood rhythm is an intriguing finding worthy of further investigation. A cross-cultural study of the effects of different social interaction patterns on mood might be illuminating. For example, Mediterranean cultures typically socialise later in the evening and have different morning-evening orientations (e.g., Smith et al. 2002a), so how do their mood profiles compare? Interestingly, Roberts' data on evening faces suggests that Mediterranean mood troughs might coincide with afternoon siestas. The faces and the other self-experiments on sleep, mood, and health also point to the value of conducting an epidemiological study to examine whether health outcomes in different jobs relate to exposure to light, faces, and standing at work.

In relation to self-experimentation as a method for generating ideas, Roberts makes the important point that there is a lack of methods for idea generation in science. One unmentioned method is communication of ideas between researchers, because ideas from one researcher can be a useful stimulus for others. To this end, scientific journals are an important vehicle for communicating mature ideas, but they are not suited or designed for the communication of speculative ideas. It is surely no coincidence that Roberts assembled 10 self-experiments for a single publication; each experiment on its own would probably have been considered too preliminary by a prestigious journal. The behavioral sciences might therefore benefit from a journal or journal section devoted to the publication of new ideas. Submissions would still need to be judged stringently, but greater emphasis could be put on novelty, insight, and likely utility, and less emphasis put on robustness and comprehensiveness. However, there is a danger that published results would be viewed as equivalent to those from more rigorous studies. Applying this point to the present case, for example, I am concerned that people might adopt Roberts' methods for weight control before they have been vouched valid and safe by well-controlled studies using population samples.

My final point concerns the ethics of self-experimentation. The history of self-experimentation (Altman 1987/1998) suggests that researchers are willing to put themselves at greater risk in pursuit of their goals than they would, or could, put others. Although individuals should probably have greater rights over their own level of risk than that of others, ethical codes should ensure that researchers take advice on their self-experiments and provide protection from potential external pressures to self-experiment.

ACKNOWLEDGMENTS

The support of the Economic and Social Research Council (ESRC) U.K. is gratefully acknowledged. The work was part of the programme of the ESRC Centre for Organisation and Innovation.

The birth of a confounded idea: The joys and pitfalls of self-experimentation

Martin Voracek[a] and Maryanne L. Fisher[b]

[a]*School of Psychology, University of Vienna, A-1010 Vienna, Austria;* [b]*Department of Psychology, St. Mary's University, Halifax, Nova Scotia, B3H 3C3, Canada.* **martin.voracek@univie.ac.at** **mlfisher@smu.ca**
http://mailbox.univie.ac.at/martin.voracek/

Abstract: According to Roberts, self-experimentation is a viable tool for idea generation in the behavioral sciences. Here we discuss some limitations of this assertion, as well as particular design and data-analytic shortcomings of his experiments.

Roberts has a fresh approach to experimental methodology. His contribution is extremely enjoyable to read, as it contains many innovative ideas as well as personal autobiography of an avid experimentalist. It is rare to read articles discussing the personal nature of sleep, mood, health, and weight, especially in the fashion done by Roberts. Rather than focus on the content of the experiments and the subsequent theory development, here we will discuss two aspects of the target article: (1) the applicability of self-experimentation for generating new research ideas, and (2) particular design and data-analytic shortcomings of Roberts' experiments.

Self-experimentation for the generation of new ideas. According to Roberts, the generation of research ideas has been overlooked in discussions of scientific methodology. He proposes self-experimentation as a useful way of filling this omission. However, idea generation is not necessarily a neglected topic. Research articles regularly include directions for future work which typically lead to new ideas and studies. As well, qualitative research is frequently used in psychology as a way of generating informed hypotheses for a given topic.

Further, self-experimentation is a questionable method for generating ideas for many fields within psychology, excluding cognition and visual perception. For example, it is not possible to use self-experimentation to explore cross-cultural factors or to study individual differences. Roberts argues that self-experimentation should be incorporated into scientific methodology, but it is clear that until it can be generalized to a wide variety of disciplines, self-experimentation's usefulness is diminished.

The precursors used to support the value of Roberts' self-experimentation seem to have been cumbersomely researched, and include no major or recent findings. The rarity of the approach is hardly surprising, given the limited application of any findings. Most of the inclusions are medical, and the application of self-experimentation to psychological issues beyond cognition and perception remains unclear.

Design and data-analytic shortcomings of self-experimentation. We have several concerns with regards to design issues. First, our hesitation in accepting the presented experiments surrounds the potential for confounding factors and the resulting erroneously inferred causality. Roberts seems to have performed multiple experiments simultaneously, and therefore, the conclusions are suspect. This problem is compounded by the fact that the author must have had expectancy effects in spite of his claims to the contrary. Using daily changes in habits as his method of experimentation, Roberts would be hard-pressed not to be vigilant for daily changes in mood, sleep, eating habits, and so forth. Blinded experimentation is a useful way of eliminating bias, but self-experimentation does not allow for any form of blinded procedure. Second, we are perturbed with the use of the self-rating scales as a viable measure of mood. The presented scales appear to have poor reliability and lack several methodological factors such as external validation, accuracy, inter-observer reliability, and scale delineation (Spector 1992). As for the findings, many of them are explained in a post hoc manner, which we argue is not considered good practice. Third, the generality of the findings remains unproven, as none of the effects have been replicated in a second individual. One should not forget that there is a large history of errors through self-inspection and

self-experimentation – most notably from psychoanalysis, which started with Freud's self-analysis.

We also have several statistical concerns. Roberts claims that single-subject experiments do not require different inferential statistical analysis than typical, multiple-subject experimental and observational research. We object to this claim and approach to the analysis. It is contradictory that Roberts favors visual display of data and exploratory data analysis, citing Tukey's (1977) seminal work, and at the same time reports numerous inferential statistical tests. On the other hand, the target article is silent about effect size indicators of treatments or interventions (Cohen 1988). Effect sizes are critical for exploratory data analysis and should be included. Also, the overuse of the loess procedure, locally weighted regression (Cleveland 1993), is problematic as the data are often noisy and may lead to dubious conclusions.

Further, there is a serious omission concerning the smoothing of time-series data, and appropriate time-series analyses are never performed. The loess procedure is not appropriate for time-series data which are necessarily serially correlated. Specifically, the loess procedure is a regression technique for serially uncorrelated data with independent realizations of a random variable.

Roberts' contribution represents 10 years of self-experimentation, and yet we remain uninformed about its reliability, confounds, generality, interindividual differences, and causality. Unfortunately, this approach is comprised solely of single-subject observations over time, and thus tells us nothing about the variability among persons that is the core of psychology. Nevertheless, the target article does incorporate one extremely useful reminder: Data is an encyclopedia for future research.

Introspection and intuition in the decision sciences

Daniel John Zizzo

Department of Economics, University of Oxford, Oxford, OX1 3UQ, United Kingdom. **daniel.zizzo@economics.ox.ac.uk**
http://www.econ.ox.ac.uk/Research/Ree/ZizzoWebpage.htm

Abstract: Self-experimentation is uncommon in the decision sciences, but mental experiments are common; for example, intuition and introspection are often used by theoretical economists as justifications for their models. While introspection can be useful for the generation of ideas, it can also be overused and become a comfortable illusion for the theorist and an obstacle for science.

This commentary complements the target article by analyzing the role that self-experimentation and mental experiments play in research on decision-making. Whereas Roberts appears to claim that participation in one's own experiments is common in human experimental psychology, my expectation is that, as a rule, papers in the decision sciences fail to have the experimenters as part of the sample. Of course, exceptions exist, especially in consumer research (e.g., Earl 2001), although even then they are controversial and usually thought of in terms of researcher introspection (Gould 1995; Wallendorf & Brucks 1993). They also tend to be less systematic than Roberts' own work (e.g., Earl [2001] is just a case study).

Roberts' target article makes sufficiently clear that *self-experimentation* means to run an experiment with oneself as a subject, going through one or more experimental conditions on the basis of a carefully drawn experimental design. In the decision sciences, the researcher would gather behavioral data, such as those one could get from other subjects, but presumably in a more flexible way (one of the advantages mentioned by Roberts); she may also collect nonbehavioral data, the reliability of which lies in introspection, and which may be richer with herself as a subject. While self-experimentation is uncommon, *mental experiments* are common and even pervasive: the experimenter considers what choices she (or some other economic agents) would make in a hypothetical situation. Mental experiments "work" insofar as introspection works. As recognized by Hutchison (1938), they have a role to play in the generation of scientific ideas (Earl 2001), which fits well with the target article. Introspection was used to defend the concept of cardinal utility in the first half of the twentieth century (Lewin 1996). Even today, economists all too often refer to "intuition" and "reasonableness" as the great virtues of their theoretical models of decision-making, thus calling for their readers, and students, to rely on their introspection (e.g., Varian 1999, p. 3).

Unfortunately, introspection has limits. First, people differ in cognitive abilities and social preferences, and population heterogeneity cannot be captured in an experiment with size $n = 1$. Second, introspection may be marred by various judgement biases, such as confirmation and hindsight bias (see Camerer 1995). For example, Dawes and Mulford (1997) showed that conjunction effects in memory recollection may explain some purported evidence for alien kidnappings (conjunction effects occur whenever an agent ranks the conjunction of two events as more likely than the less likely of the two events: Stolarz-Fantino et al. 2003). Third, there is a dissociation between what people know and learn explicitly (e.g., the laws of physics enabling people to ride bikes) and what they can actually do and learn implicitly (e.g., actually riding a bike; see Shanks & St. John 1994). For example, Zizzo (2003) found a dissociation between verbal and behavioral learning in a conjunction effect task: verbal responses were sensitive, but actual behavioral choices were entirely insensitive to the amount of verbal instructions being provided. Introspection may give access to the first source of knowledge and learning, but not to the second. Introspection may give no answers, or the wrong answers, where implicit knowledge and learning are important; for example, in the conjunction effect task example, it may lead a researcher to think that with the right feedback the effect may disappear, even though it would not in relation to behavioral choices.

Fourth, the potential dependence of behavior on *explicit* knowledge is *also* a problem. Assume that a decision scientist said "people never commit conjunction fallacies" and as evidence she brought the fact that she does not commit them. The problem is obvious: the evidence for *P* is weak if the agent can cause *P* because of her beliefs about *P.* Elster (1983) made an important distinction between primary and secondary states. Primary states are what an agent can directly choose (one can go to bed at 10 p.m.); secondary states may occur as a by-product of other states, including primary states, but are not directly chosen (an agent may go to bed at 10 p.m., and this may help her get to sleep soon afterwards, but she cannot *choose* to fall asleep at exactly 10 p.m.). For secondary states, while concerns about expectations and placebo effects do exist, the link they have with the experimental variable being investigated is indirect. Conversely, in the case of primary states, a researcher directly chooses the value of the experimental variable: for example, she chooses what action to take in a social dilemma, or which lottery to play out if she is asked to pick one of two. This creates difficulties. First, whether consciously or not, having a theory about *P* may modify a researcher's behavior to be more congruent with *P.* More important, to the extent that *P* is ruled by one's own explicit knowledge, *P will* be acted out if *P* has normative value, for example, the researcher believes that it is rational for an agent to do *P.* One of the founders of expected utility theory, Savage (1954), did commit the so-called Allais paradox, thus disproving his own theory; but, when this was pointed out to him, he said that, on second thought, the normative appeal of expected utility theory was stronger and that he would now revise his choice accordingly. In another example, economists were very surprised that their intuition about behavior in so-called ultimatum games did not fit people's behavior (Roth 1995), even though non-economists would have thought the modal choice as the intuitive one. Because of their training, decision scientists may find it all too easy to consider, by introspection, their normative theories as having descriptive value for "normal" decision-makers, but that may only be because their intuition has

been honed and changed by years of training. Introspection can thus be overused to become a comfortable illusion for the theorist, and an obstacle for science. Experiments with oneself as subject may be easier in the decision sciences as implied by Roberts, but they are also of more limited value.

Author's Response

Self-experimentation: Friend or foe?

Seth Roberts

Department of Psychology, University of California at Berkeley, Berkeley, CA 94720-1650. **roberts@socrates.berkeley.edu**

Abstract: The topics discussed in this response are in four broad areas: (1) *Idea generation,* including the failure to discuss and teach idea generation and how to nurture new ideas (sect. R2), sources of ideas worth testing with self-experimentation (sect. R3), and unusual features of the situation that may have increased the discovery rate (sect. R4); (2) *Miscellaneous methodological issues,* such as the value of mental experiments (sect. R5) and the limitations of double-blind experiments (sect. R6); (3) *Subject-matter issues,* including the relationship of this work to other evolutionary psychology studies and a new way to test evolutionary explanations (sect. R7), as well as questions about the mood results (sect. R8) and the weight results (sect. R9); and (4) *Self-experimentation,* including its difficulty (sect. R10) and future (sect. R11).

R1. Introduction

Because self-experimentation is unfamiliar, commentators on the target article resemble early adopters, the first to consider buying new products, in their willingness to take it seriously. I am grateful for their thoughtful remarks.

R2. Missing traditions

The target article begins by noting the absence of data-gathering methods designed to generate ideas (see "Missing methods") but **Glenn** and **Cabanac** point out that this is part of a larger gap. Glenn saw in the target article "a sometimes subtle but pervasive message that our understanding of scientific method is limited by the grip of traditional formulations." One of those traditions is not discussing the origin of ideas. Glenn notes how, when she was a graduate student, discussions of scientific method began *after* the hypothesis had been formulated. This state of affairs has persisted; there is little discussion of idea generation in recent textbooks (e.g., Cozby 2004; Stangor 2004) and scientific articles that introduce a new idea rarely explain its source (Feynman 1972). Moreover, as Cabanac says, granting agencies have no place in their budgets for idea generation. Their attitude seems to be: *If it happens, fine, but don't expect us to pay for it.* This tradition, as Glenn says, creates missed opportunities – discoveries that could have been made but were not.

A closely related tradition is no discussion of how to nurture new ideas. If you have a new idea, what should you do? Test it, sure, but how? The target-article research made me vividly aware that I could not answer this question nor find anything written about it. Trial and error led to one answer: *Take baby steps,* that is, do the simplest, easiest test that could increase or decrease the plausibility of the new idea (Roberts 2001). In other words, raise the bar slowly. The larger the step – the larger the difference between what you plan to do and what you have already done – the more untested assumptions you are making. The more untested assumptions, the more likely one of them is wrong. This might seem obvious, but most research proposals I encounter involving new ideas do not abide by it. The result, as **Glenn** says, is a lot of wasted effort. A few years ago, I attended a research meeting at a leading school of public health. The eight or ten persons at the meeting (almost all professors or persons with Ph.D.s) were deciding how to do a study of asthma. They were about to begin a large and expensive experiment, involving about 100 families, to measure the effect of an apartment cleaning on the incidence of asthma and allergies. They had done a pilot study with two or three families; the next step was the study itself. I suggested that they first do a larger pilot study, pointing out that the existing pilot study had not tested several assumptions that the larger study was based on. No one agreed with this suggestion, and two of the persons with doctorates said they disagreed with it. The study went ahead with no additional pilot work and was a disaster. Recruitment of families turned out to be far more difficult than expected. Funding had come from the National Institutes of Health, supporting **Cabanac**'s doubts about the wisdom of granting agencies.

R3. Sour\ces of ideas to test

The target article might have been titled "Self-experimentation as a source of ideas for conventional research" or "Self-experimentation as a growth medium for new ideas," because, in each example, self-experimentation made a new idea more plausible. **Lubart & Mouchiroud** wonder where the nurtured ideas came from. "Being able to test ideas cheaply [via self-experimentation] allows one to entertain a wider set of ideas," they write. "But these ideas need to be generated."

How were these ideas generated? For each idea that led to or was supported by self-experimentation, Table R1 gives the precipitating event (what caused me to think of it) and states whether written records helped generate it. The line between *experiment* (done to see what happens) and *natural experiment* (a change that happens for other reasons) is blurry, because actions may have more than one goal. I ate less-processed food to lose weight (Example 1), but I also did it to find out if what I had been telling my students, based on research with rats, was true for humans. Table R1 calls it a natural experiment, because my main goal was to lose weight. Table R1 distinguishes between *short* experiments (lasting a few days or less) and *long* experiments (lasting longer), because a short experiment may be nothing more than trying something new.

Table R1 supports several commentators' views about how to generate ideas worth testing. **Schegloff** believes that "sustained and systematic observation" is under-appreciated. Indeed, Table R1 shows that written records were helpful in three or four cases. **Lubart & Mouchi-**

Table R1. *Sources of tested ideas*

Example	New idea	Precipitating event	Records helpful?
1	Eating less-processed food causes weight loss	Published research (Sclafani & Springer 1976)	No
1	Weight loss reduces sleep duration	Natural experiment (change of diet to lose weight)	Yes
1	Water-rich diet causes weight loss (*not confirmed*)	Listening (student's experience)	No
1	Breakfast affects early awakening	Short experiment (change of breakfast)	Yes
2	Watching TV has the same effect as human contact on circadian oscillator	Published research (Szalai 1972)	No
2	Morning TV influences mood	Short experiment (morning TV)	No
2	Evening faces influence mood	Natural experiment (dinner with friends)	No
2	East-west travel eliminates faces effect	Natural experiment (trip to Europe)	No
3	Standing causes weight loss (*not confirmed*)	Listening (many hours/day of walking associated with weight loss)	No
3	Standing improves sleep	Long experiment (standing more)	Probably
4	Morning light improves mood	Listening (friend's experience)	No
4	Morning light improves sleep	Short experiment (outside walks)	Yes
5	Better sleep can eliminate colds	Long experiment (more standing and morning light)	No
6	Water causes weight loss	Listening (weight loss on unusual diet)	No
7	Weight control theory	Published research (Ramirez 1990a)	No
7	Low-glycemic-index food causes weight loss	Prediction of theory	No
8	Low-glycemic-index pasta causes weight loss (*not confirmed*)	Published research (list of glycemic indices)	No
9	Sushi causes weight loss	Natural experiment (all-you-can-eat sushi meal)	No
10	Sugar water causes weight loss	Natural experiment (soft drinks in Paris)	No

Note. Experiment = Self-experiment. Short = Lasted a few days or less. Long = Lasted more than a few days. Records helpful? = Did written records of mine help inspire the idea? *Not confirmed* = Self-experiment made idea less plausible.

roud emphasize taking advantage of "natural experiments." Natural experiments led to five ideas. **Moore & Sellen** say that you should "simply talk to people," that is, listen to them. Listening generated four ideas. Short experiments led to three ideas, whereas long experiments led to two ideas. Lubart & Mouchiroud are right that self-experimentation's biggest use was as an easy way to test new ideas; it was just one of several sources of ideas to test.

R4. What else mattered?

According to the target article, the rate of discovery of new cause–effect relationships was high because the work combined (a) self-experimentation and (b) new or unexploited theories. The theories narrowed the possibilities. Self-experimentation made it possible to test many of the remaining possibilities. Because a self-experiment measures many things at once, these tests often generated new ideas.

Several commentators suggest that other features shared by the examples helped raise the rate of discovery. For example, **Lubart & Mouchiroud** suggest that I was unusually motivated because I was studying "problems that [I] considered to be personally important" (para. 5). One way to determine what else mattered is to examine other discoveries. The discovery that the n-3 fatty acids found in fish are heart-protective is a convenient comparison, because Burr (2000) used it to derive the following seven rules about scientific progress.

1. "*Seek out populations that have unusual disease patterns and unusual diets and visit them*" (Burr 2000, p. 397S, italics added). Hugh Sinclair, a British physiologist, visited the Eskimos and noticed their lack of heart disease in spite of a diet very high in fat. This agrees with **Schegloff**'s belief in the value of observation and **Halberg, Cornélissen & Schack**'s (**Halberg et al.**'s) and **Grote**'s views that science and everyday life (including travel) can be profitably mixed.

2. "*Look out for natural experiments*" (Burr 2000, p. 397S), a principle which **Lubart & Mouchiroud** mention. During World War II, there was a sharp change in the Norwegian diet (less meat, more fish) and a sharp decrease in heart disease at the same time. This is more support for **Schegloff**'s belief in observation.

3. "*Use data that were originally collected for another purpose*" (Burr 2000, p. 397S). Data collected for another purpose were used to test the hypothesis that fish consumption reduced heart disease. Likewise, data collected for another purpose were useful in Examples 1–5 of the target article.

4. "*Look out for observations that do not fit in with the received wisdom*" (Burr 2000, p. 397S). When Sinclair visited the Eskimos, all fat was supposed to promote heart disease. This too supports **Schegloff**'s view of the value of observation. All of the target article examples, except perhaps 4 (morning light) and 5 (colds), contain observations that did not fit with received wisdom.

5. "*Look outside your field of research for ideas*" (Burr 2000, p. 398S). **Lubart & Mouchiroud** and **Miller** make similar points about the value of contact with other people and other areas of knowledge. The target article strongly supports this view. In Example 1 (breakfast), an animal-behavior fact (anticipatory activity) helped explain something about sleep. In Example 2 (faces), a sociological study (Sza-

lai 1972) suggested a circadian-rhythm experiment. Examples 7–10 show how Pavlovian-conditioning research led to new ideas about weight control. **Glenn** notes that I used operant-conditioning designs (e.g., long baselines) to study other topics.

6. *"If you can, be a subject in your own study."* Sinclair put himself on a diet of seal meat and fish for 100 days in 1976 and achieved record bleeding times as a result (Burr 2000, p. 398S, from Holman 1991). This helped to show that the Eskimo diet made a difference.

7. *"Research can be enjoyable"* (Burr 2000, p. 398S). I think Burr means that you should try to combine research and pleasure (such as travel to exotic places). This is not far from **Lubart & Mouchiroud**'s point that one reason for the success of my self-experimentation was its personal value.

So Burr (2000), discussing an unrelated discovery, agrees with several of the commentators' points. Perhaps the clearest agreement is about the value of: (a) Contact with other people and other areas of knowledge, (b) Observation, and (c) Personal involvement and/or enjoyment.

R5. What's worse than self-experimentation?

Zizzo writes well about a type of evidence that some psychologists question even more than self-experimentation namely, *mental experiments*, whose basic datum is what the data gatherer says he would do or feel in a certain situation. To many experimental psychologists, the hierarchy of data goes like this (from best to worse):

1. Property of action. Reaction time or percent correct, for example.

2. Verbal report of an internal state (**Miller**'s "private events"). Rated mood or certainty, for example.

3. Verbal report of a hypothetical action or hypothetical internal state. Statement of what one would do with lottery winnings, for example. Many decision-making studies gather this type of data (Thaler 1987).

With each type of evidence one can (a) study oneself (worse) or other people (better), (b) study one person (worse) or several people (better), and (c) make one measurement (worse) or multiple measurements (better) per subject – a total of four dimensions. Most experimental-psychology research is on the "better" side of all four: It measures actions, studies other people, studies several people, and makes multiple measurements per subject. Mental experiments are usually on the "worse" side of all four. The target-article research falls in between these extremes.

Zizzo's point, that even mental experiments – the lowest of the low – have value, agrees with common sense. We do mental experiments every day. For example: "How would I feel if I ate that?" (I am in a café). Their outcomes guide our actions. Plainly they are more accurate than chance. To treat them as worthless is to lose information that might be useful. In a subtle way, Zizzo is saying about behavioral science what Jacobs (2000) said about economies, that a society that treats certain people as outcasts loses the ability to have their work (e.g., toilet cleaning) be the seeds of economic growth (e.g., a cleaning business). Just as ostracism can slow economic growth, demands for methodological "correctness" can slow scientific progress. Over the last 20 to 30 years, few new treatments for bipolar disorder have been developed. Post and Luckenbaugh (2003) attributed this to too-high standards of evidence. "Many of us in the academic community," they wrote, "have inadvertently participated in the limitation of a generation of research on bipolar illness . . . in part by demands for methodological purity or study comprehensiveness that can rarely be achieved" (Post & Luckenbaugh 2003, p. 71).

R6. Miscellaneous methodology

Booth recommends double-blind studies. **Voracek & Fisher** recommend some sort of blinding (to eliminate bias). However, none of the target-article treatments, like many medical treatments, allows such experiments. Double-blind experiments are still used to obtain Food and Drug Administration approval of new drugs, but, as far as I can tell, they are being replaced in the research literature by experiments that compare the new treatment to one or more accepted treatments. An example is Jenkins et al. (2003), a study that measured the effect of three treatments (two accepted, one new) on cholesterol: a low-fat diet, a low-fat diet plus lovastatin, and a low-fat diet plus several foods that individually lower cholesterol.

Double-blind experiments have serious limitations. These include: recruitment difficulty (because potential subjects have a 50% chance of receiving a worthless treatment); compliance difficulty (why take a pill that does nothing?); trouble maintaining ignorance about the treatment (when the new treatment has noticeable side effects, as powerful treatments often do); the difficulty of implementing blinding and correcting mistakes in the implementation; failure to reproduce the conditions under which the new treatment would be used, if it works (if used clinically, recipients would know what they are getting); and failure to answer the question of most interest to clinicians and patients: What works best? Post and Luckenbaugh (2003) made similar criticisms. The new style of experimentation has none of these limitations. No design is perfect, but, as evidence increases that most placebos have little effect (Hrobjartsson & Gotzsche 2001; Kienle & Kiene 1997), double-blind experiments become even less desirable. Examples 6–10, taken together, are a single-subject version of the new style of experimentation: Five ways of losing weight, all believed and hoped to work, were compared. One of the diets (eating low-glycemic-index food), although new at the time, is now common.

Voracek & Fisher disagree with what they describe as my claim that "single-subject experiments do not require different inferential statistical analysis than typical, multiple-subject experimental, and observational research." Unfortunately, they do not say why. As Tukey (1969) emphasized, all experiments are $n = 1$ in many ways. One of their concerns is serial correlation ("appropriate time-series analyses are never performed"). It would be more accurate to say that they are not reported. As far as I could tell, serial correlations were too small to matter. In some cases, the data are given in enough detail so that this conclusion can be checked. When they are not, I am happy to supply the data to interested readers. As with $n = 1$, the issue is not restricted to self-experimentation. Few human experiments study all subjects at the same time. Rather, one person is tested on Monday, another on Wednesday, and so on. The time-series aspect is almost always ignored in the published analyses. Voracek & Fisher also say "the loess procedure is not appropriate for time-series data which are necessarily serially correlated." The statistical tests based on the loess

fits shown in Figures 14 and 15 assume uncorrelated residuals. In those figures, the abscissa is not time. The loess fits to time series (e.g., Fig. 1) do not assume zero serial correlation. They are just a way to draw a line.

According to **Voracek & Fisher,** "it is contradictory that Roberts favors visual display of data and exploratory data analysis, citing Tukey's (1977) seminal work, and at the same time reports numerous inferential statistical tests." Tukey (1980) emphasized the need for both sorts of analyses.

Voracek & Fisher wonder about confoundings: "Roberts seems to have performed multiple experiments simultaneously, and therefore, the conclusions are suspect." Unfortunately, they do not give examples. Dates given with the figures can be used to establish when various changes were made and thereby reassure oneself that one experiment did not interfere with another.

Miller says that I "produced no methodological breakthroughs in measuring" mood and how rested I felt. Quite true. Economy of means (one subject, simple designs, simple measurements, low cost, low tech, no lab) is no breakthrough, but it is a virtue.

R7. Evolutionary psychology

Miller asks: Is the target article evolutionary psychology? **Glenn** notes that many of the core ideas are evolutionary, but that it differs from previous evolutionary psychology because "the experimental method [is brought] to bear on hypotheses heretofore examined only in terms of correlational research."

What is evolutionary psychology? Buss et al. (1998) provided a definition by listing "thirty recent examples of empirical discoveries about humans generated by thinking about adaptation and selection" (p. 544). The target article's Examples 2 (faces) and 3 (standing) were generated by thinking about evolution. The key concept, however, was not adaptation or selection, but mismatch, the idea that our way of life can change much faster than our genes. An evolutionary idea was some of the initial support for the weight-control theory proposed in Example 7, and that theory led to Examples 7–10.

The target article differs from Buss et al.'s (1998) "thirty recent examples" in several ways:

1. As **Glenn** says, it is experimental. Almost all of the Buss et al. examples are correlational.

2. The theory of Examples 1–5 assumes maladaptation, that what happens now (e.g., early awakening) is *not* what happened in the Stone Age. The theories behind all of the Buss et al. examples seem to assume the opposite: What we currently observe is more or less what happened in the Stone Age.

3. Examples 2 and 3 and the weight-control theory have obvious practical applications, unlike any of the Buss et al. examples.

So the target article offers new support for evolutionary theorizing.

Questioning my evolutionary explanation of Example 2 (faces), **Totterdell** asks, "Why is a mood rhythm required to synchronize mood when there are many other more immediate unconscious and conscious processes known to bring about mood synchrony in groups (Kelly & Barsade 2001)?" My argument that the mood rhythm developed because it helped synchronize mood does not assume this rhythm was needed to synchronize mood; it just assumes such synchrony was beneficial.

Different mechanisms with the same effect are common, even ubiquitous, in biology. The hedonic shift noticed by **Cabanac** (water that usually felt unpleasantly cold felt good when he was hot) helps cool us when we are hot; so does the fact that we sweat more when hot. Many enzymes repair DNA (Sancar & Sancar 1988). Multiple mechanisms help us adjust to dim light (Woodhouse & Campbell 1975). We locate sounds using more than one mechanism (Middlebrooks & Green 1991). Because multiple mechanisms with the same effect are common, they can be used to assess evolutionary explanations. In the examples just given, the outcomes (constant body temperature, DNA repair, and so on) are obviously beneficial. If there are multiple mechanisms that produce Outcome X, it is more plausible that X was good (beneficial under Stone-Age conditions) than that X was bad. So the existence of multiple mechanisms that produce X can be used to argue that X was good. For example, suppose women prefer rich husbands to poor ones. An evolutionary explanation is that this preference is genetic and evolved because it gave the children of women with such genes more resources – supposedly a good thing. The assumption that more resources for one's children is good becomes more plausible if there are other mechanisms with the same effect (more resources for one's children). It becomes less plausible if other mechanisms have the opposite effect (less resources for one's children). According to this reasoning, the existence of other mechanisms that synchronize mood (Kelly & Barsade 2001; Totterdell et al. 1998) *supports* my assumption that the synchrony was beneficial. As far as I know, this is a new way to judge evolutionary explanations (see Buss et al. 1998 and Andrews et al. 2002 for the usual ways) – for which we can thank **Totterdell.**

R8. Mood

Totterdell is perhaps the ideal commentator on the mood results. He has studied mood with great sophistication (e.g., Totterdell et al. 1994), and his observations of his infant daughter (Totterdell 2001) resemble self-experimentation in several ways. "Both morning and evening faces," writes Totterdell, "caused a trough in mood about 18 hours after the stimulus" making it "plausible that the mood oscillation was dependent only on the social zeitgeber rather than being gated by a light-sensitive clock." One reason for assuming gating by a light-sensitive clock is that changing the timing of morning faces by just one hour in either direction diminished their effect (target article Fig. 8, upper panel; Roberts & Neuringer 1998). Faces in the late morning and afternoon had no effect, something I noticed many times but did not show in an experiment. More evidence for the involvement of a light-sensitive clock is that travel across time zones temporarily eliminated the effect (target article Fig. 9).

Totterdell wonders whether "seeing faces during Stone-Age sexual activity [was] also confined to the early morning." I suspect, but did not show, that seeing a face very close has little effect. Viewing distance clearly matters (Fig. 7; Roberts & Neuringer 1998).

Totterdell points out that "models based on the joint action of a circadian clock and a sleep need process have also

been very successful in predicting levels of alert mood." I measured not alertness but happiness, serenity, and eagerness. Alertness is about noticing stimuli; eagerness is about initiating action. As far as I can tell, I was quite alert almost the entire time I was awake, while my eagerness varied greatly (target article Fig. 5, lower panel).

I think I understand **Moore & Sellen**'s proposal: The very first morning I felt unusually happy (the morning of the accidental discovery), my mood was high due to chance. I was happy on subsequent days due to the (false) belief that I had discovered something. It is unclear to me how this idea explains other findings, such as the biphasic shape of the effect (Fig. 5, lower panel).

Moore & Sellen would presumably predict that the effect would not be repeatable with anyone else because it would not be a discovery. Other commentators have other doubts. **Booth** believes repetitions of the basic effect are "contaminated by . . . [my] knowledge of previous observations." **Totterdell** wonders if the effect depends on my "abstinence from evening interaction" and "sleep difficulties." **Voracek & Fisher** say "there is a large history of errors through self-inspection and self-experimentation" and wonder about generality and individual differences. For information about replications of the basic results, see http://psychology.berkeley.edu/pdf/roberts-replications.pdf.

R9. Weight

Booth believes that the set-point idea is wrong, or "redundant." He prefers to think of opposing negative feedback functions, which is what I think of when I try to imagine how a set point is produced. For more about the set-point concept, see Cabanac and Gosselin (1996) and their 26 references, both pro and con. The broader problem for Booth is that, in spite of what he considers my mistakes, I found a very effective and counterintuitive way of losing weight (Example 10). About this, all he says is "a lot of fructose without glucose is poorly absorbed and the resulting upset could reduce hunger." I do not recall any upset stomachs, and the daily amount of fructose was neither large nor consumed all at once. Booth knows so much about food, eating, and appetite, I can only assume his heart wasn't in thinking of an alternative explanation of Example 10.

Miller and **Voracek & Fisher** question the generality of the results. Miller wonders if the striking effect of fructose water (Example 10) depended on my earlier weight-loss experiences (Examples 1, 6–9) or other unusual aspects of my life. Moreover, according to Voracek & Miller, "there is a large history of errors through self-inspection and self-experimentation." After the target article was written, I learned of several rat studies where sugar (sucrose) water caused great weight loss (Hamilton 1971; Hamilton & Timmons 1976; Hamilton et al. 1980; Sclafani 1973). For information about the experience of those who have tried the diet, see http://psychology.berkeley.edu/pdf/roberts-replications.pdf.

One person's experience is worth describing here because it makes a larger point. A *BBS* reviewer tried the sugar-water diet and noticed that, at a dose that was "considerably lower" than my lowest dose, the fructose caused "pain in joints that were ordinarily pain-free." The reviewer had recorded joint pain for 20 years and had found that "distinctive patterns of pain accompany certain additives in food." The pain went away when each additive was avoided and returned within hours when it was consumed again. "Using a standard design of fructose, no-fructose, etc., in a matter of two weeks it was clearly demonstrable that I have a strong sensitivity to fructose (at least in the powdered form I used)," the reviewer wrote, which was "quite surprising." The reviewer concluded that self-experimentation by "an informed citizenry," plus aggregation of the results, similar to what **Halberg et al.** do with blood pressure, "may be the only way for us to ever recognize the full range of differences among individuals on health and psychological dimensions."

R10. How difficult is self-experimentation?

At least two of the commentators believe that self-experimentation similar to what the target article describes is very difficult. "The number of readers who are prepared to commit themselves to such a path in the future," writes **Schegloff,** "is surely limited, to say the least." **Miller** says the same thing obliquely when he kindly states that the work "may even constitute heroic achievement" and that some parts "inspire awe."

Perhaps it *was* heroic to drink 5 liters of water per day (Example 6). Aside from Example 6, however, none of it seemed like much trouble. Almost all the treatments took advantage of what was going to happen anyway. In Example 1 (breakfast), I was going to eat breakfast anyway; I just varied what I ate. I was going to sleep anyway; the sleep measurements took only a few seconds per night. For Example 2 (faces), I watched the faces while I walked on a treadmill, which was exercise; I was going to exercise anyway. The average reader probably cringes at the thought of standing 8 hours per day (Examples 3 and 5), but I was already going to do several things, such as work and watch TV, that could be done standing rather than sitting if I changed the environment (e.g., raised the TV). After my legs became stronger, standing 8 hrs/day was not painful. Fluorescent light exposure (Examples 4 and 5) happened while I walked on a treadmill watching TV, that is, during exercise. Eating lots of sushi (Example 9) was cumbersome and expensive but pleasant as well. The target-article research can be seen as a blend of previous self-experimentation, which was more lab-like (e.g., Ebbinghaus 1885/1913), and epidemiology, which takes advantage of preexisting differences.

On the other hand, most of the target-article treatments were more difficult than what is asked of most experimental subjects. Standing 8 hrs/day (Examples 3 and 5), drinking 5 liters of water per day (Example 6), and eating diets of pasta (Example 8) or sushi (Example 9) for weeks are considerably more difficult tasks than what subjects do in a typical sleep or weight experiment. **Lubart & Mouchiroud** correctly say that the problems I studied were "personally important," so I was "highly motivated to solve [them]." But it was more than that. A big reason why I persisted, and did relatively difficult things, was that prior self-experimentation had been successful. Success of the weight-loss treatment of Example 1 (eating less-processed food) motivated Examples 6–10. That skipping breakfast in Example 1 reduced early awakening motivated Examples 2 (morning faces), 3 (standing), and 4 (morning light). The short-term successes of Examples 3 and 4 motivated the

long-term changes and record-keeping of Example 5. Compared to the benefits, in other words, the research was not hard at all. I slept much better (Examples 1, 3, and 4). My health greatly improved (Example 5). My mood improved (Example 2). With little suffering, I reached a weight I considered ideal (Examples 1, 6–10). I went from eating three meals a day to eating one a day, saving about 2 hrs/day. In some ways, the large benefit/cost ratio is the most surprising feature of this work.

Bernstein (1997) experienced a similar benefit/cost ratio – a lifetime of much better health and freedom from diabetic complications, at the cost of about a year of not-difficult self-experimentation. In both cases, Bernstein's and mine, earlier work by others was crucial to our success. Bernstein could not have done what he did without the discovery of insulin, not to mention the blood-glucose meter he used and basic research in nutrition and physiology; and in a similar way, my work depended on the creative work of many people. Example 2 (faces) could not have happened without Wever (1979) and Szalai (1972). Example 10 (sugar water) could not have happened without Ramirez (1990a), not to mention Pavlov, Sclafani, and Cabanac. In Bernstein's case and mine, it is hard to avoid concluding that our self-experimentation (small, quick, flexible) had a large benefit/cost ratio because it found the last piece of a puzzle, a piece that (big, slow, inflexible) conventional research could not easily find.

R11. The market for self-experimentation

According to Christensen (1997), advances in technology have often caused the leading companies within an industry to lose their lead. It happened several times in the disk drive industry as drives became smaller. The most profitable companies when most drives were 14 inches in diameter lost their lead when 8-inch drives became popular. Several folded. The same thing happened when 5.25-inch drives were introduced and again when 3.5-inch drives came along. In the mechanical-shovel industry, starting around World War II, there was a shift from cable-driven shovels to hydraulic ones. Few of the firms that made cable-driven shovels survived the transition. In these and other examples, Christensen identified a common thread: The industry-leading firms were "held captive by their customers" (p. 18). They failed to appreciate the new technology because the new technology did not – in the beginning – do what their customers asked for. At first, the capacity of 8-inch drives was inadequate for the mainframes owned by the buyers of 14-inch drives. When they began, makers of 8-inch drives (and 5.25-inch drives and 3.5-inch drives) had to find *new* customers for disk drives, such as makers of minicomputers, desktop computers, and laptop computers. As the smaller drives improved (as the capacity of 8-inch drives increased, for example), they came to dominate the market previously occupied by the larger drives. When hydraulically driven mechanical shovels were first introduced, they were too small for most construction jobs. They were sold as tractor attachments; they did very small jobs previously done by hand, such as digging ditches from water lines in the street to the foundations of new houses. To the makers of big shovels, digging tiny ditches seemed like a trivial market. They saw no threat. But the maximum size of hydraulic shovels slowly increased. Eventually they could dig much larger holes and this spelled doom for the makers of cable-driven shovels. In these and other examples, the new product was seriously deficient when it first appeared. The first 8-inch disk drives did not store enough data. The first hydraulic shovels held too little dirt. But with the help of new customers, who wanted different features than the customers for the older product, the new product survived and improved and eventually replaced the older product. Christensen concluded it is unwise to dismiss a new version of a familiar product just because you (or your customers) do not want it. Others may want it. With experience, it may improve.

Self-experimentation, it is safe to say, is considered of little or no importance by the vast majority of behavioral and biomedical scientists (e.g., Jasienski 1996). Judging by publications, it is very rare. It has serious limitations compared to more traditional experimentation. The commentators sum up these limitations: the possible influence of expectations and desires (**Booth, Voracek & Fisher**), lack of information about generality across persons (**Miller,** Voracek & Fisher), and a limited range of application (Voracek & Fisher). But if self-experimentation can find even a small market, then experience can lead to improvement.

For many years, the scientific uses of self-experimentation were mainly medical. Doctors tried dangerous or difficult treatments on themselves (Altman 1987/1998). **Booth**'s half-pound-chocolate lunch is in that tradition. To an experimental psychologist, one person doing something once to see what happens is hardly science (just as digging tiny ditches is hardly construction work). Science or not, it persisted. Surely it will never be common, any prominent doctor or scientist would have said. But in 1969, as the target article mentions, Richard Bernstein, concerned about his own diabetes, obtained a new blood-glucose meter that required only one drop of blood. He used it to intensively study the effect of insulin and diet on his own blood glucose levels. This is closer to an experimental psychologist's idea of "good science," that is, there was a lot of repetition and careful study of cause and effect. Of course, it could be dismissed as therapy, not science, with results helpful only to Bernstein. But the second half of this dismissal would be wrong. His work led to the current use of blood-glucose meters by millions of diabetics, who practice a simple form of self-experimentation when they use meter readings to adjust their diet and insulin use. I know four people who use these meters. One said it was the greatest advance in diabetes treatment since the discovery of insulin. Few researchers have helped more people than Bernstein. **Voracek & Fisher**'s diabetic friends would not agree that the precursors I cite "include no major or recent findings" and that the results of self-experimentation have "limited application."

Christensen (1997) distinguished two ways that sales can expand: *downmarket* and *upmarket*. Going downmarket means finding entirely new customers for the product, who are inevitably less profitable and prestigious than the old customers, at least initially. For example, selling tractor attachments was less profitable and prestigious than selling giant shovels; selling 8-inch disk drives to minicomputer users was initially less profitable and prestigious than selling 14-inch disk drives to computer centers. Going upmarket means capturing customers from the older version of the product (e.g., when a cable-driven shovel is replaced by a hydraulic one). Going downmarket always happens first, according to Christensen; experience gained with the new customers helps penetrate the older market. Bernstein took

self-experimentation downmarket. He found entirely new users of experimentation (persons with diabetes). This had no effect on the usual users (scientists), but it greatly reduced doctor's visits, just as hydraulic-shovel tractor attachments had no effect on big contractors, but greatly reduced muscle-powered ditch-digging. One indication of the lack of profit and prestige in what Bernstein did is that hardly anyone has heard of him. The target article takes self-experimentation upmarket, in the sense that it is published in a mainstream journal (displacing an article based on traditional methods) and deals with mainstream scientific questions.

Totterdell and **Glicksohn** propose ways that self-experimentation might move further upmarket (further replacing more conventional science). Totterdell suggests that journal space be devoted to "new ideas," which would probably mean more space for self-experimentation. Glicksohn shows how self-experimentation can be used to answer important methodological questions; in his particular case, the optimal epoch size for analysis of EEG data. His example is from an experiment with two subjects rather than one, but that hardly matters. His broad point is that an experiment with one subject can help you do a later experiment with several subjects. The first experiment (with $n = 1$) can choose a procedural parameter or method of analysis; it can also generate a prediction. In these cases, to present only the second experiment (with $n > 1$) is to substantially understate the evidence. Any result is more impressive if predicted, but the basis of the prediction needs to be spelled out.

Halberg et al. and **Grote** propose ways that self-experimentation could move downmarket (into lower-prestige areas where many scientists would not deign to go). The first disk drives were found mostly at large companies and colleges – places with computer centers. Now they are in millions of places, an expansion made possible by smaller disk drives. Halberg et al. and Grote suggest the expansion of experimentation (or at least science-like observation) into the lives of millions of non-scientists. The expansion is possible because of the small "size" of self-experimentation. Halberg et al. emphasize prognosis. They show that long-term ambulatory measurement of blood pressure, done outside of medical settings, has predictive value. It is a small step from using such measurements for prognosis to using them to guide treatment. Grote emphasizes treatment. She extends what Bernstein did to more behavioral problems, especially those involving self-discipline, such as smoking cessation. To prevent relapse, Grote says, it may help to give individuals more control over their treatment. Just as a diabetic adjusts diet and insulin dose based on blood-glucose readings, so someone who is trying to stop smoking might adjust her activities based on self-measurement of her desire to smoke. No one believes that one size fits all in the treatment of diabetes or, probably, anything else; Grote argues, as Bernstein successfully argued, that a lot of the fitting is best done by the person being fit.

History is on the side of **Halberg et al.** and **Grote.** Several scientific activities have spread into the general public: (1) *Literature searches.* Non-scientists make heavy use of Medline, the National Library of Medicine's online retrieval system (Cline & Haynes 2001). (2) *Result sharing* (emphasized by **Totterdell**). Support groups and Internet discussion groups for diseases and other problems include considerable sharing of experience – how well a certain treatment worked, for instance. (3) *Sophisticated and systematic measurement.* Blood glucose meters, blood pressure meters, and cholesterol meters are available to the general public. A *BBS* reviewer of the target article mentioned BirdSource, described on its website (*http://www.birdsource.org*) as "a revolutionary partnership between citizens and scientists." BirdSource allows bird-watchers in different parts of North America to combine their observations to learn about bird migration, population, and conservation. All three activities (literature searching, result sharing, and sophisticated and systematic measurement) are growing, as far as I can tell; for example, Medline searches by the general public have greatly increased in recent years (Cline & Haynes 2001), the price of cholesterol meters is decreasing, and BirdSource began around 1997.

Another step in the direction of more self-experimentation (and possibly another example of the "first downmarket, then upmarket" pattern) is the growth of *n-of-1 trials,* which are randomized controlled clinical trials with just one patient. The original idea was that clinicians would use such trials to find the best treatment and dosage for a patient (Guyatt et al. 1990; Johannessen et al. 1990), an improvement over simple trial and error (which is how I originally viewed self-experimentation – as an improvement over simple trial and error). *N*-of-1 trials were, among other things, a way to follow-up large randomized controlled trials, which "all too often . . . are carried out on populations so ill-characterized that clinicians cannot be sure whether it is appropriate to extrapolate results to the individual patient confronting them" (Price & Evans 2002, p. 227). More recently, though, *n*-of-1 trials have been proposed to *replace* large randomized controlled trials (Post & Luckenbaugh 2003).

Human nature is also on the side of **Halberg et al.** and **Grote.** Before science was a profession, it was a hobby, which means some people enjoy it for its own sake (and which BirdSource takes advantage of). If a hobby has tangible benefits, such as lower blood pressure or reduced risk of relapse, so much stronger the motivation to do it.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Allen, R. P. (1991) Early morning awakening insomnia: Bright-light treatment. In: *Case studies in insomnia,* ed. P. J. Hauri. Plenum Medical. [aSR]

Altman, L. K. (1987/1998) *Who goes first? The story of self-experimentation in medicine.* Random House/University of California Press. (Random House edition, 1987; University of California Press edition, 1998). [IG, arSR, PT]

American Psychiatric Association. (1994) *Diagnostic and statistical manual of mental disorders,* 4th edition. [*DSM-IV*] [aSR]

Anderson, J. W., Story, L. J., Zettwoch, N. C., Gustafson, N. J. & Jefferson, B. S. (1989) Metabolic effects of fructose supplementation in diabetic individuals. *Diabetes Care* 12:337–44. [aSR]

Andrews, P. W., Gangestad, S. W. & Matthews, D. (2002) Adaptationism – how to carry out an exaptionist program. *Behavioral and Brain Sciences* 25:489–504. [rSR]

Apfelbaum, M., Reinberg, A., Nillus, P. & Halberg, F. (1969) Rythmes circadiens de l'alternance veille-sommeil pendant l'isolement souterrain de sept jeune femmes. *La Presse Medicale* 77:879–82. [aSR]

Baer, D. M. (1998) Self-experimentation: Some judgmental comments on self-experimentation. *Self-Experimentation Self-Control Communications* 10:1–5. [aSR]

Barcroft, J. & Verzár, F. (1931) The effect of exposure to cold on the pulse rate and respiration of man. *Journal of Physiology* 71:373–80. [MC]

Bartlett, F. (1958) *Thinking: An experimental and social study.* Basic Books. [aSR]

Benca, R. M., Obermeyer, W. H., Thisted, R. A. & Gillin, J. C. (1992) Sleep and psychiatric disorders. *Archives of General Psychiatry* 49:651–68. [aSR]

Benca, R. M. & Quintans, J. (1997) Sleep and host defenses: A review. *Sleep* 20:1027–37. [aSR]

Bernstein, I. L. (1980) Effects of interference stimuli on the acquisition of learned aversions to foods in the rat. *Journal of Comparative and Physiological Psychology* 94:921–31. [aSR]

Bernstein, R. K. (1997) *Dr. Bernstein's diabetes solution: A complete guide to achieving normal blood sugars.* Little, Brown. [arSR]

Beveridge, W. I. B. (1957) *The art of scientific investigation.* W. W. Norton. [aSR]

Bixler, E. O., Vgontzas, A. N., Lin, H.-M., Vela-Bueno, A. & Kales, A. (2002) Insomnia in Central Pennsylvania. *Journal of Psychosomatic Research* 53:589–92. [aSR]

Blair, A. J., Booth, D. A., Lewis, V. J. & Wainwright, C. J. (1989) The relative success of official and informal weight reduction techniques. *Psychology and Health* 3:195–206. [DAB]

Blair, S. N. & Connelly, J. C. (1996) How much physical activity should we do? The case for moderate amounts and intensities of physical activity. *Research Quarterly for Exercise and Sport* 67:193–205. [aSR]

Blair, S. N., Kohl, H. W., Gordon, N. F. & Paffenbarger, R. S., Jr. (1992) How much physical activity is good for health? *Annual Review of Public Health* 13:99–126. [aSR]

Blood, M. L., Sack, R. L., Percy, D. C. & Pen, J. C. (1997) A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. *Sleep* 20:388–95. [aSR]

Boivin, D. B., Czeisler, C. A., Dijk, D.-J., Duffy, J. F., Folkard, S., Minors, D. S., Totterdell, P. & Waterhouse, J. M. (1997) Complex interaction of the sleep-wake cycle and circadian phase modulates mood in healthy subjects. *Archives of General Psychiatry* 54:145–52. [aSR, PT]

Boivin, D. B., Duffy, J. F., Kronauer, R. E. & Czeisler, C. A. (1996) Dose-response relationships for resetting of human circadian clock by light. *Nature* 379:540–42. [aSR]

Bolles, R. C. (1990) A functionalistic approach to feeding. In: *Taste, experience, and feeding,* ed. E. D. Capaldi & T. L. Powley. American Psychological Association. [aSR]

Bolles, R. C. & Stokes, L. W. (1965) Rat's anticipation of diurnal and adiurnal feeding. *Journal of Comparative and Physiological Psychology* 60:290–94. [aSR]

Booth, D. A. (1989) The effect of dietary starches and sugars on satiety and on mental state and performance. In: *Dietary starches and sugars in Man: A comparison,* ed. J. Dobbing, pp. 225–49. Springer Verlag. [DAB]

(1990) How not to think about immediate dietary and postingestional influences on appetites and satieties (Commentary on Ramirez 1990). *Appetite* 14:171–79. [DAB]

(1998) Promoting culture supportive of the least fattening patterns of eating, drinking, and physical activity. (Prevention of Obesity: Berzelius Symposium and Satellite of 8th International Congress of Obesity). *Appetite* 31:417–20. [DAB]

Booth, D. A., Chase, A. & Campbell, A. T. (1970) Relative effectiveness of protein in the late stages of appetite suppression in man. *Physiology and Behavior* 5:1299–302. [DAB]

Booth, D. A. & Freeman, R. P. J. (1993) Discriminative measurement of feature integration in object recognition. *Acta Psychologica* 84:1–16. [DAB]

Booth, D. A., Knibb, R. C., Platts, R. G., Armstrong, A. M., Macdonald, A. & Booth, I. W. (1999) The psychology of perceived food intolerance. *Leatherhead Food RA Food Industry Journal* 2(3):325–34. [DAB]

Booth, D. A., Mobini, S., Earl, T. & Wainwright, C. J. (2003) Consumer-specified instrumental quality of short-dough cookie texture using penetrometry and break force. *Journal of Food Science: Sensory and Nutritive Qualities of Food* 68(1):382–87. [DAB]

Booth, D. A. & Smith, E. B. O. (1964) Urinary excretion of some purine bases in normal and schizophrenic subjects. *British Journal of Psychiatry* 110:582–87. [DAB]

Booth, D. A., Thompson, A. L. & Shahedian, B. (1983) A robust, brief measure of an individual's most preferred level of salt in an ordinary foodstuff. *Appetite* 4:301–12. [DAB]

Booth, D. A., Toates, F. M. & Platt, S. V. (1976) Control system for hunger and its implications in animals and man. In: *Hunger: Basic mechanisms and clinical implications,* ed. D. Novin, W. Wyrwicka & G. A. Bray, pp. 127–42. Raven Press. [DAB]

Borbely, A. A. & Achermann, P. (1992) Concepts and models of sleep regulation: an overview. *Journal of Sleep Research* 1:63–79. [aSR]

Boulos, Z., Rosenwasser, A. M. & Terman, M. (1980) Feeding schedules and the circadian organization of behavior in the rat. *Behavioral and Brain Research* 1:39–65. [aSR]

Boulos, Z. & Terman, M. (1980) Food availability and daily biological rhythms. *Neuroscience and Biobehavioral Reviews* 4:119–31. [aSR]

Box, G. E. P., Hunter, W. G. & Hunter, J. S. (1978) *Statistics for experimenters: An introduction to design, data analysis, and model building.* Wiley. [aSR]

Boyden, S. V. (1970) *The impact of civilisation on the biology of man.* Australian National University Press. [aSR]

Brand-Miller, J. C., Holt, S. H. A., Pawlak, D. B. & McMillan, J. (2002) Glycemic index and obesity. *American Journal of Clinical Nutrition* 76:281S–285S. [aSR]

Bravata, D. M., Sanders, L., Huang, J., Krumholz, H. M., Olkin, I., Gardner, C. D. & Bravata, D. M. (2003) Efficacy and safety of low-carbohydrate diets. A systematic review. *Journal of the American Medical Association* 289:1837–950. [DAB]

Breslau, N., Roth, T., Rosenthal, L. & Andreski, P. (1996) Sleep disturbance and psychiatric disorders: A longitudinal epidemiological study of young adults. *Biological Psychiatry* 39:411–18. [aSR]

Brown, G. W. (1998) Loss and depressive disorders. In: *Adversity, stress, and psychopathology,* ed. B. P. Dohrenwend. Oxford University Press. [aSR]

Brown, G. W. & Harris, T. (1978) *Social origins of depression.* Free Press. [aSR]

Brown, K. S. (1995) Testing the most curious subject – oneself. *The Scientist* 9(1):10. [aSR]

Brownell, K. D. (1987) *The LEARN program for weight control.* University of Pennsylvania School of Medicine. [aSR]

Browning, W. D. (1997) Boosting productivity with IEQ improvements. *Building Design and Construction,* April:50–52. [aSR]

Bunney, W. E. & Bunney, B. G. (2000) Molecular clock genes in man and lower animals: Possible implications for circadian abnormalities in depression. *Neuropsychopharmacology* 22:335–45. [aSR]

Burkett, L. (1987) Smoking Cessation. *Self-Experimentation Self-Control Communication.* Available at: http://www.people.ukans.edu/~grote [IG]

Burr, M. L. (2000) Lessons from the story of n-3 fatty acids. *American Journal of Clinical Nutrition* 71:397S–398S. [rSR]

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L. & Wakefield, J. C. (1998) Adaptations, exaptations, and spandrels. *American Psychologist* 53:533–48. [rSR]

Cabanac, M. (1995) *La quête du plaisir.* Editions Liber. [MC, aSR]

Cabanac, M. & Gosselin, C. (1996) Ponderostat: Hoarding behavior satisfies the condition for a lipostat in the rat. *Appetite* 27:251–61. [rSR]

Cabanac, M. & Morrissette, J. (1992) Acute, but not chronic, exercise lowers the body weight set-point in male rats. *Physiology and Behavior* 52:1173–77. [aSR]

Cabanac, M. & Rabe, E. F. (1976) Influence of a monotonous food on body weight regulation in humans. *Physiology and Behavior* 17:675–78. [aSR]

Camerer, C. (1995) Individual decision making. In: *The handbook of experimental economics,* ed. J. H. Kagel & A. E. Roth. Princeton University Press. [DJZ]

Campbell, D. T. (1960) Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Review* 67:380–400. [aSR]

Campbell, S. S., Eastman, C. I., Terman, M., Lewy, A. J., Boulos, Z. & Dijk, D. J. (1995) Light treatment for sleep disorders: Consensus report. I. Chronology of seminal studies in humans. *Journal of Biological Rhythms* 10:105–109. [aSR]

Campbell, T. C. & Junshi, C. (1994) Diet and chronic degenerative diseases: Perspectives from China. *American Journal of Clinical Nutrition* 59:1153S–1161S. [aSR]

Capaldi, E. D. (1996) Conditioned food preferences. In: *Why we eat what we eat,* ed. E. D. Capaldi. American Psychological Association. [aSR]

Carpenter, K. J., Harper, A. E., & Olson, R. E. (1997). Experiments that changed nutritional thinking. *Journal of Nutrition* 127:1017S–1053S. [aSR]

Carpenter, L. L., Kupfer, D. J. & Frank, E. (1986) Is diurnal variation a meaningful symptom in unipolar depression? *Journal of Affective Disorders* 11:255–64. [aSR]

Centers for Disease Control (1997) Smoking-attributable mortality and years of potential life lost – United States, 1984. *Morbidity and Mortality Weekly Report* 46:444–51. [IG]

Chagnon, N. A. (1983) *Yanomanö: The fierce people.* Holt, Rinehart and Winston. [aSR]

Chang, P. P., Ford, D. E., Mead, L. A., Cooper-Patrick, L. & Klag, M. J. (1997) Insomnia in young men and subsequent depression: The John Hopkins Precursors Study. *American Journal of Epidemiology* 146:105–14. [aSR]

Checkley, S. (1989) The relationship between biological rhythms and the affective disorders. In: *Biological rhythms in clinical practice,* ed. J. Arendt, D. S. Minors & J. M. Waterhouse. Wright. [aSR]

Christensen, C. M. (1997) *The innovator's dilemma: When new technologies cause great firms to fail.* Harvard Business School Press. [rSR]

Cleveland, W. S. (1993) *Visualizing data.* AT&T Bell Laboratories. [MV]

Cline, R. J. W. & Haynes, K. M. (2001) Consumer health information seeking on the Internet: The state of the art. *Health Education Research* 16:671–92. [rSR]

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences,* 2nd edition. Erlbaum. [MV]

Cohen, M. N. (1989) *Health and the rise of civilization.* Yale University Press. [aSR]

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S. & Gwaltney, J. M., Jr. (1997) Social ties and susceptibility to the common cold. *Journal of the American Medical Association* 277:1940–44. [aSR]

Conner, M. T., Haddon, A. V., Pickering, E. S. & Booth, D. A. (1988) Sweet tooth demonstrated: Individual differences in preference for both sweet foods and foods highly sweetened. *Journal of Applied Psychology* 73:275–80. [DAB]

Cordain, L., Brand-Miller, J., Eaton, S. B., Mann, N., Holt, S. H. H. & Speth, J. D. (2000) Plant-animal subsistence rations and macronutrient energy estimations in worldwide hunter-gatherers. *American Journal of Clinical Nutrition* 71:682–92. [aSR]

Cornélissen, G., Halberg, F., Hawkins, D., Otsuka, K. & Henke, W (1997) Individual assessment of antihypertensive response by self-starting cumulative sums. *Journal of Medical Engineering and Technology* 21:111–20. [aSR]

Cornélissen, G., Halberg, F., Schwartzkopff, O., Delmore, P., Katinas, G., Hunter, D., Tarquini, B., Tarquini, R., Perfetto, F., Watanabe, Y. & Otsuka, K. (1999) Chronomes, time structures, for chronobioengineering for "a full life." *Biomedical Instrumentation and Technology* 33:152–87. [aSR]

Cornélissen, G., Halberg, F., Wendt, H. W., Bingham, C., Sothern, R. B., Haus, E., Kleitman, E., Kleitman, N., Revilla, M. A., Revilla, M., Jr., Breus, T. K., Pimenov, K., Grigoriev, A. E., Mitish, M. D., Yatsyk, G. V. & Syutkina, E. V. (1996) Resonance of about-weekly human heart rate rhythm with solar activity change. *Biologia* (Bratislava) 51:749–56. [FH]

Corruccini, R. S. (1999) *How anthropology informs the orthodontic diagnosis of malocclusion's causes.* Edwin Mellen Press. [aSR]

Corruccini, R. S. & Kaul, S. S. (1983) The epidemiology transition and anthropology of minor chronic non-infectious diseases. *Medical Anthropology* 7:36–50. [aSR]

Cox, D. R. (1958) *Planning of experiments.* Wiley. [JG]

Cozby, P. C. (2004) *Methods in behavioral research,* 8th edition. McGraw-Hill. [rSR]

Crisp, A. H. & Stonehill, E. (1971) Aspects of the relationship between psychiatric status, sleep, nocturnal motility and nutrition. *Journal of Psychosomatic Research* 15:501–509. [aSR]

(1973) Aspects of the relationship between sleep and nutrition: A study of 375 psychiatric out-patients. *British Journal of Psychiatry* 122:379–94. [aSR]

Crisp, A. H., Stonehill, E., Fenton, G. W. & Fenwick, P. B. C. (1973) Sleep patterns in obese patients during weight reduction. *Psychotherapy and Psychosomatics* 22:159–65. [aSR]

Crosbie, J. (1995) Interrupted time-series analysis with short series: Why it is problematic; how it can be improved. In: *The analysis of change,* ed. J. M. Gottman, pp. 361–95. Erlbaum. [JG]

Curb, J. D. & Marcus, E. B. (1991) Body fat and obesity in Japanese Americans. *American Journal of Clinical Nutrition* 53:1552S–1155S. [aSR]

Czeisler, C. A. (1995) The effect of light on the human circadian pacemaker. *Ciba Foundation Symposium* 183:254–90. [aSR]

Czeisler, C. A., Kronauer, R. E., Johnson, M. P., Allan, J. S., Johnson, T. S. & Dumont, M. (1989) Action of light on the human circadian pacemaker: Treatment of patients with circadian rhythm sleep disorders. In: *Sleep '88,* ed. J. Horne. Gustave Fischer Verlag. [aSR]

Danesh, J., Collins, R. & Peto, R. (1997) Chronic infections and coronary heart disease: Is there a link? *Lancet* 350:430–36. [aSR]

Dawes, R. M. & Mulford M. (1997) Diagnoses of alien kidnappings that result from conjunction effects in memory. In: *The UFO Invasion: The Roswell Incident, Alien Abductions, and Government Coverups,* ed. K. Frazier, B. Karr & J. Nickell. Prometheus Press. [DJZ]

Descartes, R. (1637/1931) Discourse on the method of rightly conducting the reason and seeking for truth in the sciences. In: *The Philosophical Works of Descartes, vol. 1,* trans. E. S. Haldane & G. R. T. Ross. Cambridge University Press, 1931. [SCM]

Diagnostic and Statistical Manual of Mental Disorders (1994) 4th edition. [*DSM-IV*] American Psychiatric Association. [aSR]

Diamond, H. & Diamond, M. (1985) *Fit for life.* Warner Books. [aSR]

Diamond, J. (1992) Diabetes running wild. *Nature* 357:362–63. [aSR]

Dinan, T. G. (1999) The physical consequence of depressive illness. *British Medical Journal* 318:826. [aSR]

Dinges, D. F., Douglas, S. D., Hamarman, S., Zaugg, L. & Kapoor, S. (1995) Sleep deprivation and human immune function. *Advances in Neuroimmunology* 5:97–110. [aSR]

Donnison, C. P. (1937) *Civilization and disease.* Baillière, Tindall & Cox. [aSR]

Doucet, E., Imbeault, P., St-Pierre, S., Almeras, N., Mauriege, P., Richard, D. & Tremblay, A. (2000) Appetite after weight loss by energy restriction and a low-fat diet-exercise follow-up. *International Journal of Obesity and Related Metabolic Disorders* 24:906–14. [aSR]

Douglas, A. S., Allan, T. M. & Rawles, J. M. (1991) Composition of seasonality of disease. *Scottish Medical Journal* 36:76–82. [aSR]

Earl, P. E. (2001) Simon's travel theorem and the demand for live music. *Journal of Economic Psychology* 22:335–58. [DJZ]

Eaton, S. B. & Eaton, S. B., III (1999) The evolutionary context of chronic degenerative diseases. In: *Evolution in health and disease,* ed. S. C. Stearns. Oxford University Press. [aSR]

Ebbinghaus, H. (1885/1913) *Memory: A contribution to experimental psychology.* Columbia University. (Original work published 1885.) [arSR]

Egeland, J. A. & Hostetter, A. M. (1983) Amish study, I: Affective disorders among the Amish, 1976–1980. *American Journal of Psychiatry* 140:6–61. [aSR]

Ehlers, C. L., Frank, E. & Kupfer, D. J. (1988) Social zeitgebers and biological rhythms. *Archives of General Psychiatry* 45:948–52. [aSR]

Elster, J. (1983) *Sour grapes: Studies in the subversion of rationality.* Cambridge University Press. [DJZ]

Fantino, M. (1976) Effet de l'alimentation intragastrique au long cours chez l'homme. *Journal de Physiologie* 72:86A. [aSR]

Fenner, D. P. & Bernstein, I. L. (1984) Interference in food aversion conditioning by reducing drug dose or conditioning trials. *Behavioral and Neural Biology* 40:114–18. [aSR]

Feynman, R. P. (1972) The development of the space-time view of quantum electrodynamics. In: *Nobel lectures on physics, vol. 4,* pp. 155–78. Elsevier. [rSR]

Folkard, S. & Akerstedt, T. (1989) Towards the prediction of alertness on abnormal sleep/wake schedules. In: *Vigilance and performance in automatized systems,* ed. A. Coblentz. Kluwer Academic. [aSR]

Folkard, S., Akerstedt, T., Macdonald, I., Tucker, P. & Spencer, M. B. (1999) Beyond the three-process model of alertness: Estimating phase, time on shift, and successive night effects. *Journal of Biological Rhythms* 14:577–87. [PT]

Ford, D. E. & Kamerow, D. B. (1989) Epidemiologic study of sleep disturbances and psychiatric disorders: An opportunity for prevention? *Journal of the American Medical Association* 262:1479–84. [aSR]

French, J. A., Wainwright, C. J., Booth, D. A. & Hamilton, J. (1992) Effects of meat species and particle size on postprandial satiety. *Proceedings of the Nutrition Society* 51:57A. [DAB]

Furka, A. (2001) Foreword. In: *High-throughput synthesis: Principles and practices,* ed. I. Sucholeiki. Marcel Dekker. [aSR]

Gardner, J. & Oswald, A. (2002) Does money buy happiness? A longitudinal study using data on windfalls. In: *Royal Economic Society Annual Conference, 81.* Royal Economic Society. [SCM]

Garrow, J. S. & Stalley, S. (1975) Is there a "set point" for human body-weight? *Proceedings of the Nutrition Society* 34:84A–85A. [aSR]

Georgi, F. (1947) Psychophysische Korrelationen: III: Psychiatrische Probleme im Licte der Rhythmusforschung. *Schweizerische Medizinische Wochenschrift* 49:1276–80. [aSR]

Gibbs, W. W. (1996) Gaining on fat. *Scientific American* 275(2):88–94. [aSR]

Gillberg, M. (1997) Human sleep/wake regulation. *Acta Anaesthesiologica Scandinavica Supplement* 110:8–10. [aSR]

Glicksohn, J., Friedland, G. & Salach-Nachum, A. (1990–91) Systematic self-observation of entoptic phenomena and their relation to hypnagogia. *Imagination, Cognition and Personality* 10:269–78. [JG]

Glicksohn, J., Gvirtsman, D. & Offer, S. (1995–96) The compensatory nature of mood: A single-subject time-series approach. *Imagination, Cognition and Personality* 15:385–96. [JG]

Glicksohn, J. & Myslobodsky, M. S. (1993) The representation of patterns of structural brain asymmetry in normal individuals. *Neuropsychologia* 31:145–59. [JG]

Gould, S. J. (1995) Researcher introspection as a method in consumer research: Applications, issues, and implications. *Journal of Consumer Research* 21:719–22. [DJZ]

Grady, D. (2000) Trading obesity's dangers for surgery's. *New York Times,* October 12:A1. [aSR]

Grote, I. (1997) The emergence of combination therapies intensifies the need for generalization technologies. In: *Environment and behavior,* ed. D. M. Baer & E. M. Pinkston, pp. 289–99. Westview. [IG]

(1999) The behavior of the scientist: Epistemological tools science-makers use and games they play. *Behavior and Social Issues* 9:47–53.

(submitted) *Visual analysis of basic scientific design.* [IG]

Gundel, A., Polyakov, V. V. & Zulley, J. (1997) The alteration of human sleep and circadian rhythms during spaceflight. *Journal of Sleep Research* 6:1–8. [aSR]

Guyatt, G. H., Keller, J. L., Jaeschke, R., Rosenbloom, D., Adachi, J. D. & Newhouse, M. T. (1990) The *n*-of-1 randomized controlled trial: Clinical usefulness. *Annals of Internal Medicine* 112:293–99. [rSR]

Gwinup, G. (1975) Effect of exercise alone on the weight of obese women. *Archives of Internal Medicine* 135:676–80. [aSR]

Halberg, F. (1968) Physiologic considerations underlying rhythmometry, with special reference to emotional illness. In: *Cycles biologiques et psychiatrie,* ed. J. DeAjuriaguerra. Georg. [aSR]

Halberg, F., Cornélissen, G., Otsuka, K., Schwartzkopff, O., Halberg, J. & Bakken, E. E. (2001) Chronomics. *Biomedicine and Pharmacotherapy* 55 (Suppl. 1):153–90. [FH]

Halberg, F., Cornélissen, G., Otsuka, K., Watanabe, Y., Katinas, G. S., Burioka, N., Delyukov, A., Gorgo, Y., Zhao, Z. Y., Weydahl, A., Sothern, R. B., Siegelova, J., Fiser, B., Dusek, J., Syutkina, E. V., Perfetto, F., Tarquini, R., Singh, R. B., Rhees, B., Lofstrom, D., Lofstrom, P., Johnson, P. W. C., Schwartzkopff, O. & International BIOCOS Study Group (2000) Cross-spectrally coherent ~10.5- and 21-year biological and physical cycles, magnetic storms and myocardial infarctions. *Neuroendocrinology Letters* 21:233–58. [aSR]

Halberg, F., Cornélissen, G., Schack, B., Wendt, H. W., Minne, H., Sothern, R. B., Watanabe, Y., Katinas, G., Otsuka, K. & Bakken, E. E. (2003) Blood pressure self-surveillance for health also reflects 1.3-year Richardson solar wind variation: Spin-off from chronomics. *Biomedicine and Pharmacotherapy* 57(Suppl. 1):58S–76S. [FH]

Halberg, F., Engeli, M., Hamburger, C. & Hillman, D. (1965) Spectral resolution of low-frequency, small-amplitude rhythms in excreted 17-ketosteroids; probable androgen-induced circaseptan desychronization. *Acta Endocrinologica* (Suppl.) 103:5–54. [aSR]

Hallonquist, J. D., Goldberg, M. A. & Brandes, J. S. (1986) Affective disorders and circadian rhythms. *Canadian Journal of Psychiatry* 31:259–72. [aSR]

Halmos, P. R. (1975) The problem of learning to teach: I. The teaching of problem solving. *American Mathematical Monthly* 82:466–70. [aSR]

Hamilton, L. W. (1971) Starvation induced by sucrose ingestion in the rat: Partial protection by septal lesions. *Journal of Comparative and Physiological Psychology* 77:59–69. [rSR]

Hamilton, L. W. & Timmons, C. R. (1976) Sex differences in response to taste and postingestive consequences of sugar solutions. *Physiology and Behavior* 17:221–25. [rSR]

Hamilton, L. W., Timmons, C. R. & Lerner, S. M. (1980) Caloric consequences of sugar solutions: A failure to obtain gustatory learning. *American Journal of Psychology* 93:387–407. [rSR]

Hansen, V., Lund, E. & Smith-Sivertsen, T. (1998) Self-reported mental distress under the shifting daylight in the high north. *Psychological Medicine* 28:447–52. [aSR]

Harris, L. D. (1988) Edge effects and conservation of biotic diversity. *Conservation Biology* 2:330–32. [aSR]

Haug, H.-J. (1992) Prediction of sleep deprivation outcome by diurnal variation of mood. *Biological Psychiatry* 31:271–78. [aSR]

Haug, H.-J. & Fahndrich, E. (1990). Diurnal variations of mood in depressed patients in relation to severity of depression. *Journal of Affective Disorders* 19:37–41. [aSR]

Haug, H.-J. & Wirz-Justice, A. (1993) Diurnal variation of mood in depression: Important or irrelevant? *Biological Psychiatry* 34:201–203. [aSR, PT]

Heacock, P. M., Hertzler, S. R. & Wolf, B. W. (2002) Fructose prefeeding reduces the glycemic response to a high-glycemic index, starchy food in humans. *Journal of Nutrition* 132:2601–604. [aSR]

Healy, D. & Waterhouse, J. M. (1990) The circadian system and affective disorders: Clocks or rhythms? *Chronobiology International* 7:5–10. [aSR]

(1995) The circadian system and the therapeutics of the affective disorders. *Pharmacology and Therapeutics* 65:241–63. [aSR]

Healy, D. & Williams, J. M. (1988) Dysrhythmia, dysphoria, and depression: The interaction of learned helplessness and circadian dysrhythmia in the pathogenesis of depression. *Psychological Bulletin* 103:163–78. [aSR]

(1989) Moods, misattributions and mania: An interaction of biological and psychological factors in the pathogenesis of mania. *Psychiatric Developments* 7:49–70. [aSR]

Heller, R. F. & Heller, R. F. (1997) *The carbohydrate addict's lifespan program.* Penguin Putnam. [aSR]

Herbert, T. B. & Cohen, S. (1993) Depression and immunity: A meta-analytic review. *Psychological Bulletin* 113:472–86. [aSR]

Herbert, V. (1962) Experimental nutritional folate deficiency in man. *Transactions of the Association of American Physicians* 73:307–20. [aSR]

Hightower, J. M. & Moore, D. (2003) Mercury levels in high-end consumers of fish. *Environmental Health Perspectives* 111:604–608.

Hippisley-Cox, J., Fielding, K. & Pringle, M. (1998) Depression as a risk for ischemic heart disease in men: Population based case-control study. *British Medical Journal* 316:1714–19. [aSR]

Hirsch, J., Hudgins, L. C., Leibel, R. L. & Rosenbaum, M. (1998) Diet composition and energy balance in humans. *American Journal of Clinical Nutrition* 67:551S–555S. [aSR]

Ho, A., Fekete-Mackintosh, A., Resnikoff, A. M. & Grinker, J. (1978) Sleep timing changes after weight loss in the severely obese. *Sleep Research* 7:69. [aSR]

Hoch, T. A. (1987) Smoking cessation after 8 years of smoking 20–40 cigarettes/day and after several unsuccessful attempts at stopping "cold turkey." *Self-Experimentation Self-Control Communication.* Available at: http://www.people.ukans.edu/~grote. [IG]

Holman, R. T. (1991) Hugh MacDonald Sinclair 1910–1990, advocate of the essential fatty acids. In: *Health effects of omega-3 polyunsaturated fatty acids in seafoods, World Review of Nutrition and Dietetics, vol. 66,* ed. A. P. Simopoulos, R. R. Kifer, R. E. Martin & S. M. Barlow, pp. xx–xxiv. Karger. [rSR]

Hope-Simpson, R. E. (1981) The role of season in the epidemiology of influenza. *Journal of Hygiene* 86:35–47. [aSR]

Hoptman, M. J. & Davidson, R. J. (1998) Baseline EEG asymmetries and performance on neuropsychological tasks. *Neuropsychologia* 36:1343–53. [JG]

Horgan, J. (1999) *The undiscovered mind: How the human brain defies replication, medication, and explanation.* The Free Press. [aSR]

Hostetler, J. A. (1980) *Amish society,* 3rd edition. John Hopkins University Press. [aSR]

Hrobjartsson, A. & Gotzsche, P. C. (2001) Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 21:1594–1602. [arSR]

Humphrey, G. (1951) *Thinking: An introduction to its experimental psychology.* Methuen. [JG]

Hutchison, T. W. (1938) *The significance and basic postulates of economic theory.* Augustus M. Kelley. [DJZ]

Hyman, R. (1964) *The nature of psychological inquiry.* Prentice-Hall. [aSR]

Intersalt Cooperative Research Group (1988) Intersalt: An international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. *British Medical Journal* 297:319–28. [aSR]

Irons, D. (1998) *The world's best-kept diet secrets.* Sourcebooks. [aSR]

Jacobs, J. (1984) *Cities and the wealth of nations.* Random House. [aSR]

(2000) *The nature of economies.* Random House. [aSR]

Jacobs, W. W. (1990) Effects of cigarette smoking on the human heart. *Self-Experimentation Self-Control Communication.* Available at: http://www.people.ukans.edu/~grote. [IG]

Jansson, B. (1985) Geographic mappings of colorectal cancer rates: A retrospect of studies, 1974–1984. *Cancer Detection and Prevention* 8:341–48. [aSR]

Jasienski, M. (1996) Wishful thinking and the fallacy of single-subject experimentation. *The Scientist* 10(5):10. [rSR]

Jefferson, T. O. & Tyrrell, D. (2001) Antivirals for the common cold (Cochrane review). *The Cochrane Library,* 4. Update Software. [aSR]

Jenkins, D. J. A., Kendall, C. W. C., Marchie, A., Faulkner, D. A., Wong, J. M. W., de Souza, R., Eman, A., Parker, T. L., Vidgen, E., Lapsley, K. G., Trautwein, E. A., Josse, R. G., Leiter, L. A. & Connelly, P. W. (2003) Effects of a dietary portfolio of cholesterol-lowering foods vs lovastatin on serum lipids and C-reactive protein. *Journal of the American Medical Association* 290:502–10. [rSR]

Jewett, M. E., Kronauer, R. E. & Czeisler, C. A. (1991) Light-induced suppression of endogenous circadian amplitude in humans. *Nature* 350:59–62. [aSR]

Johannessen, T., Petersen, H., Kristensen, P. & Fosstvedt, D. (1990) The controlled single subject trial. *Scandinavian Journal of Primary Health Care* 9:17–21. [rSR]

Johnson, M. A., Polgar, J., Weightman, D. & Appleton, D. (1973) Data on the distribution of fibre types in thirty-six human muscles: An autopsy study. *Journal of the Neurological Sciences* 18:111–29. [aSR]

Josling, P. (2001) Preventing the common cold with a garlic supplement: A double-blind, placebo-controlled survey. *Advances in Therapy* 18:189–93. [aSR]

Kaplan, G. A., Roberts, R. E., Camacho, T. C. & Coyne, J. C. (1987) Psychosocial predictors of depression. Prospective evidence from the human population laboratory studies. *American Journal of Epidemiology* 125:206–20. [aSR]

Keesey, R. E. & Hirvonen, M. D. (1997) Body weight set-points: Determination and adjustment. *Journal of Nutrition* 127:1875S–1883S. [aSR]

Kelly, J. R. & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes [Special issue: Affect at work: Collaborations of basic and organizational research]* 86:99–130. [rSR, PT]

Kempner, W. (1944) Treatment of kidney disease and hypertensive vascular disease with rice diet. *North Carolina Medical Journal* 5:125–33. [aSR]

Kennedy, G. C. (1950) The hypothalamic control of food intake in rats. *Proceedings of the Royal Society of London,* 137B:535–48. [aSR]

Kennedy, G. J., Kelman, H. R., Thomas, C., Wisniewski, W., Metz, H. & Bijur, P. E. (1989) Hierarchy of characteristics associated with depressive symptoms in an urban elderly sample. *American Journal of Psychiatry* 146:220–25. [aSR]

Kessler, R. C. (1997) The effects of stressful life events on depression. *Annual Review of Psychology* 48:191–214. [aSR]

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshelman, S., Wittchen, H.-U. & Kendler, K. S. (1994) Lifetime and 12-month

prevalence of DSM-III-R psychiatric disorders in the United States. *Archives of General Psychiatry* 51:8–19. [aSR]
Keys, A., Broñek, J., Henschel, A., Mickelson, O. & Taylor, H. L. (1950) *The biology of human starvation.* University of Minnesota Press. [aSR]
Kienle, G. S. & Kiene, H. (1997) The powerful placebo effect: Fact or fiction? *Journal of Clinical Epidemiology* 50:1311–18. [arSR]
Kim, S. H. & Tokura, H. (1998) Visual allesthesia: Cloth color preference in the evening under the influence of different light intensities during the daytime. *Physiology and Behavior* 65:367–70. [aSR]
Klahr, D. & Simon, H. (1999) Studies of scientific discovery: Complementary approaches and convergent findings. *Psychological Bulletin* 125:524–43. [aSR]
Klein, R. (1979) Rhythms in human performance: 1½-hour oscillations in cognitive style. *Science* 204:1326–28. [aSR]
Klerman, G. L. & Weissman, M. M. (1989) Increasing rates of depression. *Journal of American Medical Association* 261:2229–35. [aSR]
Knibb, R., Booth, D. A., Armstrong, A., Platts, R., Macdonald, A. & Booth, I. W. (1999) Episodic and semantic memory in reports of food intolerance. *Applied Cognitive Psychology* 13:451–64. [DAB]
Koestler, A. (1964) *The act of creation.* Hutchinson. [aSR]
Kohsaka, M., Fukuda, N., Honma, H., Kobayashi, R., Sakakibara, S., Koyama, E., Nakano, T. & Matsubara, H. (1999) Effects of moderately bright light on subjective evaluations in healthy elderly women. *Psychiatry and Clinical Neurosciences* 53:239–41. [aSR]
Kolata, G. (2000) How the body knows to gain or lose. *New York Times,* October 17:D1 (National Edition). [aSR]
Krall, E. A. & Dawson-Hughes, B. (1994) Walking is related to bone density and rates of bone loss. *American Journal of Medicine* 96:20–26. [aSR]
Kraybill, D. B. (1989) *The riddle of Amish culture.* John Hopkins University Press. [aSR]
Kripke, D. F. (1984) Critical interval hypotheses for depression. *Chronobiology International* 1:73–80. [aSR]
Kristofferson, A. B. (1976) Low variance stimulus-response latencies: Deterministic internal delays? *Perception and Psychophysics* 20:89–100. [aSR]
(1977) A real-time criterion theory of duration discrimination. *Perception and Psychophysics* 21:105–17. [aSR]
(1980) A quantal step function in duration discrimination. *Perception and Psychophysics* 27:300–306. [aSR]
Kromhout, D., Bloemberg, B., Seidell, J. C., Nissinen, A. & Menotti, A. (2001) Physical activity and dietary fiber determine population body fat levels: The Seven Countries Study. *International Journal of Obesity* 25:301–306. [aSR]
Krueger, J. M. (1990) Somnogenic activity of immune response modifiers. *Trends in Pharmacological Sciences* 11:122–26. [aSR]
Krueger, J. M. & Karnowsky, M. L. (1987) Sleep and the immune response. *Annals of the New York Academy of Sciences* 496:510–16. [aSR]
Lacey, J. H., Crisp, A. H., Kalucy, R. S., Hartmann, M. K. & Chen, C. N. (1975) Weight gain and the sleeping electroencephalogram: Study of 10 patients with anorexia nervosa. *British Medical Journal* 4:556–58. [aSR]
Lack, L. C., Mercer, J. D. & Wright, H. (1996) Circadian rhythms of early morning awakening insomniacs. *Journal of Sleep Research* 5:211–19. [aSR]
Lack, L. & Wright, H. (1993) The effect of evening bright light in delaying the circadian rhythms and lengthening the sleep of early awakening insomniacs. *Sleep* 16:436–43. [aSR]
Lakatos, I. (1978) *The methodology of scientific research programmes.* Cambridge University Press. [aSR]
Leibel, R. L., Rosenbaum, M. & Hirsch, J. (1995) Changes in energy expenditure resulting from altered body weight. *New England Journal of Medicine* 9:621–28. [aSR]
Leppämäki, S., Partonen, T. & Lönnqvist, J. (2002) Bright-light exposure combined with physical exercise elevates mood. *Journal of Affective Disorders* 72:139–44. [aSR]
Lewin, S. B. (1996) Economics and psychology: Lessons for our own day from the early twentieth century. *Journal of Economic Literature* 34:1293–323. [DJZ]
Lewinsohn, P. M., Rohde, P., Seeley, J. R. & Fischer, S. A. (1993) Age-cohort changes in the lifetime occurrence of depression and other mental disorders. *Journal of Abnormal Psychology* 102:110–20. [aSR]
Long, S. J., Jeffcoat, A. R. & Millward, D. J. (2000) Effect of habitual dietary-protein intake on appetite and satiety. *Appetite* 35:79–88. [DAB]
Looker, A. C., Orwoll, E. S., Johnston, C. C., Jr., Lindsay, R. L., Wahner, H. W., Dunn, W. L. , Calvo, M. S., Harris, T. B. & Heyse, S. P. (1997) Prevalence of low femoral bone density in older U.S. adults from NHANES III. *Journal of Bone and Mineral Research* 12:1761–68. [aSR]
Lorch, R. F., Jr. & Myers, J. L. (1990) Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory and Cognition* 16:149–57. [JG]
Lubart, T. I. (1999) Componential models. In: *Encyclopedia of creativity, vol. 1,* ed. M. A. Runco & S. R. Pritsker, pp. 295–300. Academic Press. [TIL]
Lubart, T., Mouchiroud, C., Tordjman, S. & Zenasni, F. (2003) *Psychologie de la créativité* [*Psychology of creativity*]. Armand Colin. [TIL]
Lubow, R. E. (1973) Latent inhibition. *Psychological Bulletin* 79:398–407. [aSR]
Lucas, F. & Sclafani, A. (1996) Food deprivation increases the rat's preference for a fatty flavor over a sweet taste. *Chemical Senses* 21:169–79. [aSR]
Ludwig, D. S. (2000) Dietary glycemic index and obesity. *Journal of Nutrition* 130:280S–283S. [aSR]
Ludwig, D. S. & Eckel, R. H. (2002) The glycemic index at 20 y. *American Journal of Clinical Nutrition* 76:264S–265S. [aSR]
Ludwig, D. S., Pereira, M. A., Kroenke, C. H., Hilner, J. E., Van Horn, L., Slattery, M. L. & Jacobs, D. R., Jr. (1999) Dietary fiber, weight gain, and cardiovascular disease risk factors in young adults. *Journal of the American Medical Association* 282:1539–46. [aSR]
Mackintosh, N. J. (1974) *The psychology of animal learning.* Academic Press. [aSR]
Mayer, J. (1968) *Overweight: Causes, cost, and control.* Prentice-Hall. [aSR]
McGuire, W. J. (1997) Creative hypothesis generating in psychology: Some useful heuristics. *Annual Review of Psychology* 48:1–30. [aSR]
McMurran, M. & Sellen, J. L. (in preparation) Offenders' motivation to change: What do people who work with offenders think? [SCM]
Medawar, P. B. (1969) *Induction and intuition in scientific thought.* American Philosophical Society.
(1984) *The limits of science.* Harper & Row. [aSR]
Michel, C. & Cabanac, M. (1999) Lipectomy, body weight, and body weight set point in rats. *Physiology & Behavior* 66:473–79. [aSR]
Middlebrooks, J. C. & Green, D. M. (1991) Sound localization by human listeners. *Annual Review of Psychology* 42:135–59. [rSR]
Miller, J. B., Foster-Powell, K. & Colagiuri, S. (1996) *The G. I. factor.* Hodder & Stoughton. [aSR]
Milton, K. (2000) Hunter-gatherer diets–a different perspective. *American Journal of Clinical Nutrition* 71:665–67. [aSR]
Minors, D. S. & Waterhouse, J. M. (1986) Circadian rhythms and their mechanisms. *Experientia* 15:1–13. [aSR]
Mistlberger, R. E., Houpt, T. A. & Moore-Ede, M. C. (1990) Food-anticipatory rhythms under 24-hour schedules of limited access to single macronutrients. *Journal of Biological Rhythms* 5:35–46. [aSR]
Moffoot, A. P. R., O'Carroll, R. E., Bennie, J., Carroll, S., Dick, H., Ebmeier, K. P. & Goodwin, G. M. (1994) Diurnal variation of mood and neuropsychological function in major depression with melancholia. *Journal of Affective Disorders* 32:257–69. [aSR]
Moldofsky, H. (1995) Sleep and the immune system. *International Journal of Immunopharmacology* 17:649–54. [aSR]
Monmaney, T. (1993) Marshall's hunch. *The New Yorker,* September 20:64–72. [aSR]
Montignac, M. (1999) *Eat yourself slim.* Erica House. [aSR]
Moos, W. H. (1998) Forward. *Biotechnology and Bioengineering (Combinatorial Chemistry)* 61:5–6. [aSR]
Morita, T. & Tokura, H. (1998) The influence of different wavelengths of light on human biological rhythms. *Applied Human Science* 17:91–96. [aSR]
Morris, J. N., Heady, J. A., Raffle, P. A. B., Roberts, C. G. & Parks, J. W. (1953) Coronary heart disease and physical activity of work. *Lancet* 2:1053–57. [aSR]
Myers, D. G. (1992) *The pursuit of happiness: Who is happy, and why.* William Morrow. [aSR]
National Library of Medicine (2002) Malocclusion of teeth. *Medline Plus Medical Encyclopedia,* located at: www.nlm.nih.gov/medlineplus/ency/article/001058.htm. Accessed November 12, 2002. [aSR]
Neel, J. V. (1994) *Physician to the gene pool.* Wiley. [aSR]
Nelson, J. C. & Charney, D. S. (1981) The symptoms of major depressive illness. *American Journal of Psychiatry* 138:1–13. [aSR]
Neuringer, A. (1981) Self-experimentation: A call for change. *Behaviorism* 9:79–94. [aSR]
(1984) Melioration and self-experimentation. *Journal of the Experimental Analysis of Behavior* 42:397–406. [IG]
Nieman, D. C. (1994) Physical activity, fitness, and infection. In: *Physical activity, fitness, and health,* ed. C. Bouchard, R. J. Shepard & T. Stephens. Human Kinetics. [aSR]
Nieto, F. J. (1998) Infections and atherosclerosis: New clues from an old hypothesis. *American Journal of Epidemiology* 148:937–48. [aSR]
O'Hara, M. W., Kohout, F. J. & Wallace, R. B. (1985) Depression among the rural elderly: A study of prevalence and correlates. *Journal of Nervous and Mental Disease* 173:582–89. [aSR]
Ohayon, M. M. (2002) Epidemiology of insomnia: What we know and what we still need to learn. *Sleep Medicine Reviews* 6:97–111. [aSR]
Oliver, J. E. (1991) *The incomplete guide to the art of discovery.* Columbia University Press. [aSR]
Orr, W. C., Hoffman, H. J. & Hegge, F. W. (1974) Ultradian rhythms in extended performance. *Aerospace Medicine* 45:995–1000. [aSR]

Pappenheimer, J. R. (1983) Induction of sleep by muramyl peptides. *Journal of Physiology* 336:1–11. [aSR]
Pavia, M. R. (2001) High-throughput organic synthesis: A perspective looking back on 10 years of combinatorial chemistry. In *High-throughput synthesis: Principles and practices,* ed. I. Sucholeiki. Marcel Dekker. [aSR]
Peck, J. W. (1976) Situational determinants of the body weights defended by normal rats and rats with hypothalamic lesions. In: *Hunger: Basic mechanisms and clinical implications,* ed. D. Novin, W. Wyrwicka & G. A. Bray, pp. 297–311. Raven Press. [DAB]
Peden, B. F. & Keniston, A. H. (1990) Self-experimentation as a tool for teaching about psychology. In: *Activity handbook for the teaching of psychology,* ed. V. P. Makowsky, C. C. Sileo, L. G. Whisttemore, C. P. Landry & M. L. Skutley. American Psychological Association. [aSR]
Petry, N. M. & Simcic, F., Jr. (2002) Recent advances in the dissemination of contingency management techniques; Clinical and research perspectives. *Journal of Substance Abuse Treatment* 23:81–86. [IG]
Pettigrew, J. D. & Miller, S. M. (1998) A 'sticky' interhemispheric switch in bipolar disorder? *Proceedings of the Royal Socey of London* (Series B: Biological Science) 265:2141–48. [aSR]
Post, R. M. & Luckenbaugh, D. A. (2003) Unique design issues in clinical trials of patients with bipolar affective disorder. *Journal of Psychiatric Research* 37:61–73. [rSR]
Price, J. D. & Evans, J. G. (2002) N-of-1 randomized controlled trials ('N-of-1 trials'): Singularly useful in geriatric medicine. *Age and Ageing* 31:227–32. [rSR]
Raben, A., Vasilaras, T. H., Møller, A. C. & Astrup, A. (2002) Sucrose compared with artificial sweeteners: Different effects on ad libitum food intake and body weight after 10 weeks in overweight subjects. *American Journal of Clinical Nutrition* 76:721–29. [aSR]
Rajaratnam, S. M. W. & Redman, J. R. (1999) Social contact synchronizes free-running activity rhythms of diurnal palm squirrels. *Physiology and Behavior* 66:21–26. [aSR]
Ramirez, I. (1987) When does sucrose increase appetite and adiposity? *Appetite* 9:1–19. [aSR]
(1990a) Stimulation of energy intake and growth by saccharin in rats. *Journal of Nutrition* 120:123–33. [arSR]
(1990b) Why do sugars taste good? *Neuroscience and Biobehavioral Reviews* 14:125–34. [aSR]
Randich, A. & LoLordo, V. M. (1979) Associative and nonassociative theories of the UCS preexposure phenomenon: Implications for Pavlovian conditioning. *Psychological Bulletin* 86:523–48. [aSR]
Raynor, H. A. (2001) Dietary variety, energy regulation, and obesity. *Psychological Bulletin* 127:325–41. [aSR]
Regestein, Q. R. & Monk, T. H. (1995) Delayed sleep phase syndrome: A review of its clinical aspects. *American Journal of Psychiatry* 152:602–608. [aSR]
Remington, N. A., Fabrigar, L. R. & Visser, P. S. (2000) Reexamining the circumplex model of affect. *Journal of Personality and Social Psychology* 79:286–300. [PT]
Richardson, J. D., Paularena, K. I., Belcher, J. W. & Lazarus, A. J. (1994) Solar wind oscillations with a 1.3-year period. *Geophysical Research Letters* 21:1559–60. [FH]
Roberts, R. M. (1989) *Serendipity: Accidental discoveries in science.* Wiley. [aSR]
Roberts, S. (1994) Arising aid (U.S. Patent No. 5,327,331). United States Department of Commerce Patent and Trademark Office. [aSR]
Roberts, S. (1998) Baer (1998) on self-experimentation. *Self-Experimentation Self-Control Communication.* Available at: http://www.people.ukans.edu/~grote. [IG]
(2001) Surprises from self-experimentation: Sleep, mood, and weight. *Chance* 14(2):7–12. [arSR]
Roberts, S. & Neuringer, A. (1998) Self-experimentation. In: *Handbook of research methods in human operant behavior,* ed. K. A. Lattal & M. Perrone. Plenum. [arSR]
Roberts, S. B. (2000) High-glycemic index foods, hunger, and obesity: Is there a connection? *Nutrition Reviews* 58:163–69. [aSR]
Rolls, B. & Barnett, R. A. (2000) *Volumetrics: Feel full on fewer calories.* HarperCollins. [aSR]
Root-Bernstein, R. S. (1989) *Discovering.* Harvard University Press. [aSR]
Rosenthal, R. (1966) *Experimenter effects in behavioral research.* Appleton-Century-Crofts. [aSR]
Roth, A. E. (1995) Bargaining experiments. In: *The handbook of experimental economics,* ed. J. H. Kagel & A. E. Roth. Princeton University Press. [DJZ]
Sancar, A. & Sancar, G. B. (1988) DNA repair enzymes. *Annual Review of Biochemistry* 57:29–67. [rSR]
Savage, L. J. (1954) *The foundations of statistics.* Wiley. [DJZ]
Schacter, D. L. (1976) The hypnagogic state: A critical review of the literature. *Psychological Bulletin* 83:452–81. [JG]
Schegloff, E. A. (1996) Issues of relevance for discourse analysis: Contingency in action, interaction and co-participant context. In: *Computational and conversational discourse: Burning issues – An interdisciplinary account,* ed. E. H. Hovy & D. Scott, pp. 3–38. Springer Verlag. [EAS]
Schwartz, A. & Schwartz, R. M. (1993) *Depression: Theories and treatments.* Columbia University Press. [aSR]
Schwartz, M. W., Baskin, D. G., Kaiyala, K. J. & Woods, S. C. (1999) Model for the regulation of energy and adiposity by the central nervous system. *American Journal of Clinical Nutrition* 69:584–96. [aSR]
Schwarz, N. & Strack, F. (1999) Reports of subjective well-being: Judgmental processes and their methodological implications. In: *Well-being: The foundations of hedonic psychology,* ed. D. Kahneman, E. Diener & N. Schwarz, pp. 61–84. Russell Sage Foundation. [SCM]
Sclafani, A. (1973) Feeding inhibition and death produced by glucose ingestion in the rat. *Physiology and Behavior* 11:595–601. [rSR]
(1991) Conditioned food preferences. *Bulletin of the Psychonomic Society* 29:256–60. [aSR]
Sclafani, A., Cardieri, C., Tucker, K., Blusk, D. & Ackroff, K. (1993) Intragastric glucose but not fructose conditions robust flavor preferences in rats. *American Journal of Physiology (Regulatory, Integrative, and Comparative Physiology)* 265:R320–R325. [aSR]
Sclafani, A. & Springer, D. (1976) Dietary obesity in adult rats: Similarities to hypothalamic and human obesity syndromes. *Physiology and Behavior* 17:461–71. [arSR]
Sears, B. (1995) *The zone.* HarperCollins. [aSR]
Seidell, J. (2000) Obesity, insulin resistance, and diabetes – a worldwide epidemic. *British Journal of Nutrition* 83:S5–S8. [aSR]
Seneci, P. (2000) *Solid-phase synthesis and combinatorial technologies.* Wiley-Interscience. [aSR]
Shacham, S. (1983) A shortened version of the Profile of Mood States. *Journal of Personality Assessment* 47:305–306. [aSR]
Shanks, D. R. & St. John, M. F. (1994) Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* 17:367–447. [DJZ]
Shapiro, G. (1986) *A skeleton in the darkroom: Stories of serendipity in science.* Harper & Row. [aSR]
Shepard, R. J., Jones, G., Ishii, K., Kaneko, M. & Olbrecht, A. J. (1969) Factors affecting body density and thickness of subcutaneous fat. Data on 518 Canadian city dwellers. *American Journal of Clinical Nutrition* 22:1175–89. [aSR]
Shetty, P. S. & James, W. P. (1994) *Body mass index: A measure of chronic energy deficiency in adults.* Food and Agriculture Organization of the United Nations. [aSR]
Shoptaw, S., Rotheram-Fuller, E., Yang, X, Frosch, D., Nahom, D., Jarvic, M. E. Rawson, R. A. & Ling, W. (2002) Smoking cessation in methadone maintenance. *Addiction* 97:1317–28. [IG]
Siffre, M. (1964) *Beyond time.* McGraw-Hill. [aSR]
Simon, H. A. & Kaplan, C. A. (1989) Foundations of cognitive science. In: *Foundations of cognitive science,* ed. M. I. Posner. MIT Press. [SCM]
Simonton, D. K. (1999) *Origins of genius: Darwinian perspectives on creativity.* Oxford University Press. [aSR]
Skinner, B. F. (1976) *Walden Two.* Macmillan. [HLM]
Smith, C. S. (1981) *A search for structure: Selected essays on science, art, and history.* MIT Press. [aSR]
Smith, C. S., Folkard, S., Schmeider, R. A., Parra, L. F., Spelten, E., Almiral, H., Sen, R. N., Sahu, S., Perez, L. M. & Tisak, J. (2002a) Investigation of morning-evening orientation in six countries using the preferences scale. *Personality and Individual Differences* 32:949–68. [PT]
Smith, L. D., Best, L. A., Stubbs, A., Archibald, A. B. & Roberson-Nay, R. (2002b) Knowledge: The role of graphs and tables in hard and soft psychology. *American Psychologist* 57:749–61. [SSG]
Smith, R. (1985) Exercise and osteoporosis. *British Medical Journal* 290:1163–64. [aSR]
Sobel, R. K. (2001) Barry Marshall. *U.S. News and World Report* 131(7):39. [aSR]
Solomon, A. (2001) *The noonday demon: An atlas of depression.* Scribner. [aSR]
Soroker, N., Cohen, T., Baratz, C., Glicksohn, J. & Myslobodsky, M. S. (1994) Is there a place for ipsilesional eye patching in neglect rehabilitation? *Behavioural Neurology* 7:159–64. [JG]
Spector, P. E. (1992) *Summated rating scale construction: An introduction.* Sage. [MV]
Stangor, C. (2004) *Research methods for the behavioral sciences,* 2nd edition. Houghton Mifflin. [rSR]
Starbuck, S., Cornélissen, G. & Halberg, F. (2002) Is motivation influenced by geomagnetic activity? *Biomedicine and Pharmacotherapy* 56(Suppl. 2):289S–297S. [FH]
Sternberg, R. J. & Lubart, T. I. (1995) *Defying the crowd: Cultivating creativity in a culture of conformity.* Free Press. [TIL]
Steward, H. L., Bethea, M. C., Andrews, S. S. & Balart, L. A. (1998) *Sugar busters!* Ballantine Books.
Stolarz-Fantino, S., Fantino, E., Zizzo, D. J. & Wen, J. (2003) The conjunction

fallacy: New evidence for robustness. *American Journal of Psychology* 116:15–34. [DJZ]
Stone, A. A. (1996) Daily mood variability: Form of diurnal patterns and determinants of diurnal patterns. *Journal of Applied Social Psychology* 26:1286–1305. [aSR]
Strassmann, B. I. & Dunbar, R. I. M. (1999) Human evolution and disease: Putting the Stone Age perspective. In: *Evolution in health and disease,* ed. S. C. Stearns. Oxford University Press. [aSR]
Streblow, D. N., Orloff, S. L. & Nelson, J. A. (2001) Do pathogens accelerate atherosclerosis? *Journal of Nutrition* 131:2798S–2804S. [aSR]
Stueve, A., Dohrenwend, B. P. & Skodol, A. E. (1998) Relationships between stressful life events and episodes of major depression and nonaffective psychotic disorders: Selected results from a New York risk factor study. In: *Adversity, stress, and psychopathology,* ed. B. P. Dohrenwend. Oxford University Press. [aSR]
Suzuki, T. (1994) *The essentials of Japanese cooking.* Shufunotomo. [aSR]
Swenson, C. J., Baxter, J., Shetterly, S. M., Scarbro, S. L. & Hamman, R. F. (2000) Depressive symptoms in Hispanic and Non-Hispanic white rural elderly. *American Journal of Epidemiology* 152:1048–55. [aSR]
Szalai, A. (1972) *The use of time: Daily activities of urban and suburban populations in twelve countries.* Mouton. [arSR]
Szent-Györgyi, A. (1972) Dionysians and Apollonians. *Science* 176:966. [MC]
Takahashi, J. S., DeCoursey, P. J., Bauman, L. & Menaker, M. (1984) Spectral sensitivity of a novel photoreceptive system mediating entrainment of mammalian circadian rhythms. *Nature* 308:186–88. [aSR]
Tennesen, M. (2000) A drink to your health. *Health,* June:89–92. [aSR]
Thaler, R. (1987) The psychology of choice and the assumptions of economics. In: *Laboratory experimentation in economics: Six points of view,* ed. A. E. Roth, pp. 99–130. Cambridge University Press. [rSR]
Tordoff, M. G. & Alleva, A. M. (1990) Effect of drinking soda sweetened with aspartame or high-fructose corn syrup on food intake and body weight. *American Journal of Clinical Nutrition* 51:963–69. [aSR]
Totterdell, P. (2001) Links between time of day, sleep and unhappy mood in early infancy: An intensive case study. *British Journal of Developmental Psychology* 19:433–51. [rSR, PT]
Totterdell, P., Kellett, S., Teuchmann, K. & Briner, R. B. (1998) Evidence of mood linkage in work groups. *Journal of Personality and Social Psychology* 74:1504–15. [rSR]
Totterdell, P., Reynolds, S., Parkinson, B. & Briner, R. B. (1994) Associations of sleep with everyday mood, minor symptoms and social interaction experience. *Sleep* 17:466–75. [arSR]
Trowell, H. C. & Burkitt, D. P. (1981) *Western diseases: Their emergence and prevention.* Harvard University Press. [aSR]
Tukey, J. W. (1969) Analyzing data: Sanctification or detective work. *American Psychologist* 24:83–91. [arSR]
(1977) *Exploratory data analysis.* Addison-Wesley. [arSR, MV]
(1980) We need both exploratory and confirmatory. *American Statistician* 34:23–25. [rSR]
Turner, R. B. (1997) Epidemiology, pathogenesis, and treatment of the common cold. *Annals of Allergy, Asthma, and Immunology* 78:531–40. [aSR]
Vaillant, G. E. (1998) Natural history of male psychological health, XIV: Relationship of mood disorder vulnerability to physical health. *American Journal of Psychiatry* 155:184–91. [aSR]
Valenstein, E. S. (1998) *Blaming the brain.* The Free Press. [aSR]
Valentine, C. W. (1953) *Introduction to experimental psychology,* 5th edition. London University Tutorial Press. [DAB]
Varian, H. R. (1999) *Intermediate microeconomics: A modern approach.* Norton. [DJZ]
Vernikos-Danellis, J. & Winget, C. M. (1979) The importance of light, postural and social cues in the regulation of the plasma cortisol rhythm in man. In: *Chronopharmacology,* ed. A. Peinberg & F. Halberg. Pergamon Press. [aSR]
Vgontzas, A. N. & Kales, A. (1999) Sleep and its disorders. *Annual Review of Medicine* 50:387–400. [aSR]
Wackermann, J., Lehmann, D., Michel, C. M. & Strik, W. K. (1993) Adaptive segmentation of spontaneous EEG map series into spatially defined microstates. *International Journal of Psychophysiology* 14:269–83. [JG]
Wallendorf, M. & Brucks, M. (1993) Introspection in consumer research: Implementation and implications. *Journal of Consumer Research* 20:339–59. [DJZ]
Walsh, B. T., Goetz, R., Roose, S. P., Fingeroth, S. & Glassman, A. H. (1985) EEG-monitored sleep in anorexia nervosa and bulimia. *Biological Psychiatry* 20:947–56. [aSR]
Waterhouse, J., Minors, D., Folkard, S., Owens, D., Atkinson, G., Macdonald, I., Reilly, T., Sytnik, N. & Tucker, P. (1998) Light of domestic intensity produces phase shifts of the circadian oscillator in humans. *Neuroscience Letters* 245:97–100. [aSR]
Watford, M. (2002) Small amounts of dietary fructose dramatically increase hepatic glucose update through a novel mechanism of glucokinase activation. *Nutrition Reviews* 60:253–64. [aSR]
Webb, W. B. (1985) Sleep in industrialized settings in the northern hemisphere. *Psychological Reports* 57:591–98. [aSR]
Weerasinghe, D. P., MacIntyre, C. R. & Rubin, G. L. (2002) Seasonality of coronary artery deaths in New South Wales, Australia. *Heart* 88:30–34. [aSR]
Wehr, T. A. (1990a) Manipulations of sleep and phototherapy: Nonpharmacological alternatives in the treatment of depression. *Clinical Neuropharmacology* 13 (Suppl. 1):S54–S65. [aSR]
(1990b) Reply to Healy, D. and Waterhouse, J. M.: The circadian system and affective disorders: Clocks or rhythms? *Chronobiology International* 7:11–14. [PT]
Wehr, T. A. & Goodwin, F. K. (1983) *Circadian rhythms in psychiatry.* Boxwood Press. [aSR]
Wever, R. A. (1979) *The circadian system of man: Results of experiments under temporal isolation.* Springer-Verlag. [arSR]
Whitney, E. & Rolfes, S. (1999) *Understanding nutrition.* West/Wadsworth. [aSR]
Willett, W. C. (1998) Is dietary fat a major determinant of body fat? *American Journal of Clinical Nutrition* 67:556S–562S. [aSR]
Williams, P. T. (1997) Relationship of distance run per week to coronary disease risk factors in 8,283 male runners. *Archives of Internal Medicine* 157:191–98. [aSR]
Winfree, A. T. (1973) Resetting the amplitude of Drosophila's circadian chronometer. *Journal of Comparative Physiology* 85:105–40. [aSR]
Wing, R. R. & Hill, J. O. (2001) Successful weight loss maintenance. *Annual Review of Nutrition* 21:323–41. [aSR]
Wirtshafter, D. & Davis, J. D. (1977) Set points, settling points, and the control of body weight. *Physiology and Behavior* 19:75–78. [aSR]
Wirz-Justice, A. (1986) Light therapy for depression: Present status, problems, and perspectives. *Psychopathology* 19(Suppl. 2):136–41. [aSR]
Wolever, T. M. (1990) Relationship between dietary fiber content and composition in foods and the glycemic index. *American Journal of Clinical Nutrition* 51:72–75. [aSR]
Woodhouse, J. M. & Campbell, F. W. (1975) The role of the pupil light reflex in aiding adaptation to the dark. *Vision Research* 15:649–53. [rSR]
World Health Organization (2000) *Obesity: Preventing and managing the global epidemic.* World Health Organization. [aSR]
Zeitzer, J. M., Dijk, D.-J., Kronauer, R. E., Brown, E. N. & Czeisler, C. A. (2000) Sensitivity of the human circadian pacemaker to nocturnal light: Melatonin phase resetting and suppression. *Journal of Physiology* 526:695–702. [aSR]
Zizzo, D. J. (2003) Probability compounding in words and practice. *Theory and Decision* 54: 287–314. [DJZ]