Appendix A: Self-Experiments

Quantified Self How-To: Designing Self-Experiments[1]

By Konstantin Augemberg, Nov 14, 2012

Sometimes, you would like to know if a certain variable X in your life affects another variable, so you could use it for your advantage. Perhaps, you are curious if meditating in the evening helps to minimize stress on the following day. Or if that supplement you have been taking really helps you lose weight faster. The problem with turning to scientific studies and online reviews is that what worked for others may not necessarily work for you. The best way to find out if the cause-effect relationship between X and Y exists **for you** is to conduct *your own personal, Quantified-Self experiment*.

The first step, would be to formulate the *research hypothesis*. What is it exactly that you are trying to achieve? It should be precise and to the point, something like this: "if I do X then there will be change in Y". If I meditate in the evening, my stress levels will be lower on the following day. If I take supplement Z, then my weight loss rate will increase".

The second step is *operationalization* and selection of proper instruments and measurement procedures. Will your measurement rely on subjective tools (e.g., psychological questionnaires), more objective (e.g., electronic or balance scale, apps that measure reaction or pulse, gadgets that measure sleep quality, etc.), or both? If you rely on subjective measures, do some research to see if there are any established and scientifically sound questionnaires already exist out there?

You will also need to establish a *timeframe*. Unlike regular experiments which involve multiple participants, self-experiments have only one subject – you. Hence, in order to accumulate enough data points, self-experiments may need to last at least several days. Ultimately, it all depends on how quickly Y changes: for instance, you may capture changes in mood and sleep quality sooner than the changes in weight and body fat.

Self-Tracking Basics: Operationalization

One of the major challenges you face before you even start self-tracking is defining in plain terms how to measure the variables of interest. In experimental social sciences, there is a special term for that: operationalization. It refers to defining the variable of interest via empirical (measurable) qualities. In other words, you define it in terms of specific operations by which you will be measuring it, and then stick to that operational routine. For instance, you would like to track quality of sleep. Now, you can measure it in many different ways.

You can ask yourself upon awakening to rate your sleep on a five-point scale, from "not good at all" to "excellent". You can also count how many times you woke up during the night. Or you can record how much time does it take for you to get out of bed after you woke up. If you prefer more objective measures, you may choose to rely instead on devices like mobile phone (Sleep Cycles or Smart Alarm clock apps) or more advanced device (BodyMedia, Zeo, FitBit) to collect certain sleep statistics (e.g., length of REM

[1] http://hplusmagazine.com/2012/11/14/quantified-self-how-to-designing-self-experiments/

sleep).My word of advice: instead of relying on intuition and personal understanding of subject, turn to professionals and academics, and conduct some diligent research (e.g., using Google Scholar) before you start logging data. Trust me, no matter what you are trying to measure, the chances are that someone out there has already done some empirical research on the subject. Relying on solid scientific research is especially important if you are trying to measure a psychological trait or other intangible construct. You want to have a valid and reliable questions to measure the variable of interest, questions that have been successfully tested already on many people. Otherwise you risk to end up with messy, unreliable data that won't provide you with any meaningful insights.

Self-Tracking Basics: Minimizing Measurement Error

One of the first "reality checks" that you have to accept when starting a self-tracking project is the existence of measurement error. No matter how technologically advanced the measurement and recording tools you are using, these instruments will never be able to capture the "true value" of the object or trait that you attempt to measure. Take, for instance, body weight. If you weigh yourself several times throughout the day, and then chart the results, you will notice considerable fluctuation. Your weight will naturally go up right after breakfast and even more after lunch, and your gym scales will most likely show different results than your scales at home. These differences are caused by so called "*systematic*" error.

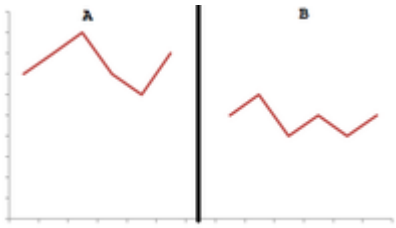
This is an error that occurs due to the misuse of instruments and changes in measurement procedure. In the above mentioned example, you weighed yourself in different time of the day, using different scales, etc. Every time you introduce a change in the measurement procedure, you affect the results. Luckily, systematic errors can almost often be eliminated or minimized by means of standardization of the measurement process. In other words, you fight systematic error by introducing the system: find the optimal conditions under which the measurement is assumed to be most accurate, and from now on stick to the same routine. Another type of error that will always follow you on your self-tracking endeavors is the "*random*" error. This is a "noise" that occurs due to uncontrollable factors and nature of the object that you are trying to measure.

I would say that the more intangible is the object, the higher is random error. For instance, latent characteristics such as emotions, mood and other psychological traits are the most difficult to capture. In this case, taking multiple measurements (or better, multiple instruments) and averaging out results could help to minimize the error. Finally, remember: in self-tracking projects, it is almost always about relative comparisons. In other words, it is not the absolute value of whatever you measure that matters, but the relative change in its values across time periods or treatments. If you keep measurement routine the same, then you will still be able to capture the change. So even if your bathroom scale constantly underestimates your weight by 1 lbs, you will still be able to see your progress, as long as you continue using the same scale.

Self-Tracking Basics: Experimental Design

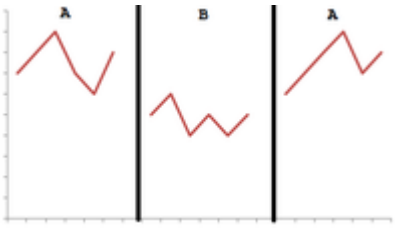
Finally, you need to choose a *design* for your experiment. Here are the three most common experimental designs for self experimentation:

(A-B) Design



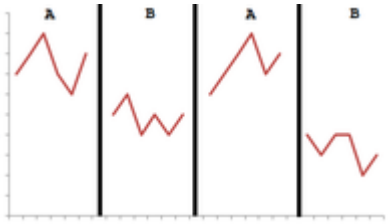
The simplest of all, this design is also considered to be the "weakest" when it comes to capturing causality. The A and B are experimental phases. Phase A is a "baseline", during which you measure Y variable under normal conditions. For instance, you spend a week or two measuring every day your weight, sleep quality, stress levels, etc. Phase B is a "treatment" phase, during which you introduce your variable X: you meditate every evening, take weight loss supplement, etc, and continue measuring Y. At the end, you compare levels of Y during both phases: did you lose more weight during the phase B? were your stress levels lower during the phase B? Based on the findings, you make conclusions about the effectiveness of the treatment.

(A-B-A) Design



This design is considered to be more advanced, as it lets you capture the changes in Y before and after the treatment. For instance, you spend the first week going to bed normally, the second week meditating before you to go sleep, and on the third week you going back to regular regimen. Looking at the changes in your stress levels across the three weeks (if there are any) helps you to conclude if the meditation works and how long does its effects last.

(A-B-A-B) Design



This design may be useful if you want to see if the intensity of treatment B is associated with intensity of the outcome. For instance, whether spending more time meditating will help you to reduce stress even more. So during the fourth phase, you may want to meditate more than once a day, or increase time of your evening meditation twofold. You then compare the changes in stress levels following both treatment phases to make the conclusion. In another version of A-B-A-B design, in either of the treatment phases, the treatment is replaced with placebo (e.g., similarly looking harmless pills instead of weight supplement; of course, during both treatment phases you don't know which supplement you are taking). The longer versions of this design, A-B-A-B-A-B and A-B1-A-B2-A-B3 (where B1, B2 and B3 are different versions of treatment, e.g., different types of meditation) are also used commonly.

Now that you have chosen the instruments and design for your experiment, you can develop a *protocol*: detailed instructions on how and when to administer treatments and take measurements. In 90% of the cases, success of experiment depends on sticking to the protocol: dedication and consistency are the key. And remember: negative results are results, too. At least you will know that that super-duper X everyone is swearing by does not work for your Y. Which means that the search for that perfect X of your own continues!

How To Self-Experiment[2]

[2] http://quantifiedself.com/2011/04/how-to-self-experiment/

Posted on April 6, 2011 by Seth Roberts

Alexandra Carmichael has asked me to write a post about how to do self-experimentation as a kind of encouragement to come to the session. Here's some of what I've learned.

1. **Easier to learn useful stuff than I expected**. In contrast to the rest of life, where things turn out harder than expected, learning useful stuff by self-experimentation was always **easier** than I expected, in the sense that the benefit/cost ratio was unexpectedly high. I learned useful things I never expected to learn. An example is acne. When I was a grad student I had acne. My dermatologist had prescribed two drugs, tetracycline and benzoyl peroxide. I believed that the tetracycline worked and the benzoyl peroxide did not work. My results showed the opposite. It hadn't occurred to me that I could be so wrong, nor that my dermatologist could be wrong (he believed both worked), nor that the establishment view (treat acne with tetracycline) could so easily be shown to be wrong.

2. **Don't be afraid of subjective measurements**. By subjective measurements I mean non-physical measurements, such as ratings of mood or how rested I felt — what professional researchers call "self-report". They routinely say self-report is misleading. At first, I wondered if my expectations and hopes would distort the measurements. As far as I can tell, that didn't happen. Instead, I found such measurements helped me learn plenty of useful stuff I couldn't have learned without it. For example, I learned how to improve my mood and how to wake up more rested.

3. **Complex experimental designs were rarely worth the extra effort**. Now and then I tried relatively complex experimental designs (e.g., randomization, a factorial experiment). Usually they were too hard.

4. **Run conditions until you get 5-40 days of flat results** (flat = what you are measuring is not going up or down). Ideal is 10-20 days. Suppose I want to compare Treatments A and B (e.g., different amounts of butter). I decide to make one measurement/day. The first step would be to do A for several days. I keep doing A until whatever I am measuring (e.g., sleep) stops steadily increasing or decreasing and then run several more days — ideally, 10-20. Then I do B for several days. I keep doing B until my measurement stops changing, then I do 10-20 more days of B. If the B measurements looked different from the A measurements, I would then return to Treatment A. It's always a good idea to run a treatment until your central measurement stops changing, and then run it longer. How much longer? I've found that less than 5 days makes me nervous. Whereas running a condition for more than 40 days of flat results is a wasted opportunity to learn more by trying a different treatment.

5. **Data analysis is easy. The most important thing is to plot measurement versus day**. It will tell you most of what you want to know. For example, most of the graphs in this paper show whatever I was measuring (sleep, weight, etc.) as a function of day.

6. **When you add data, look again at all the data.** Each time I collect new data, I plot all of the data, or at least a large chunk of it. This helps spot unexpected changes. For example, each time I measure my weight I look at a plot of my weight over the last year or so. Recently I found that cold showers caused me to gain weight, which I hadn't expected. If I hadn't looked at a year of data every time I weighed myself, it would have taken longer to notice this.

7. **Don't adjust your set**. My conclusions often contradicted expert opinion. Again and again, other data suggested my self-experimental conclusions were correct, in spite of what the experts said. Acne is one

example. Later research supported my conclusion that tetracycline didn't work. Another example is breakfast. Experts say breakfast is "the most important meal of the day." I found it caused me to wake up too early. When I stopped eating it, my sleep got better. Other data supported my conclusion. The Shangri-La Diet is a third example. According to experts, it should never work. Hundreds of stories show that it does.

The most useful lesson I learned was also the most basic. **You will be tempted to do something complicated. Don't. Do the simplest easiest thing that will tell you something**. The world was always more complicated than I realized. Gradually it sank in: Complicated (experiment) plus complicated (world) = confusion. Simple (experiment) plus complicated (world) = progress.