Best of LessWrong: July 2017

Best of LessWrong: July 2017

Subtle Forms of Confirmation Bias

There are at least two types of confirmation bias.

The first is ***selective attention:*** a tendency to pay attention to, or recall, that which confirms the hypothesis you are thinking about rather than that which speaks against it.

The second is ***selective experimentation:*** a tendency to do experiments which will confirm, rather than falsify, the hypothesis.

The standard advice for both cases seems to be "explicitly look for things which would falsify the hypothesis". I think this advice is helpful, but it is subtly wrong, especially for the selective-experimentation type of confirmation bias. Selective attention is relatively straightforward, but selective experimentation is much more complex than it initially sounds.

Looking for Falsification

What the standard (Popperian) advice tells you to do is try as hard as you can to falsify your hypothesis. You should think up experiments where your beloved hypothesis really could fail.

What this advice definitely does do is guard against the mistake of making experiments which could not falsify your hypothesis. Such a test is either violating conservation of expected evidence (by claiming to provide evidence one way without having any possibility of providing evidence the other way), or providing only very weak evidence for your claim (by looking much the same whether your claim is true or false). Looking for tests which can falsify your result steers you towards tests which would provide strong evidence, and helps you avoid violating the law of expected evidence.

However, there are more subtle ways in which confirmation bias can act.

Predicting Results in Advance

You can propose a test which would indeed fit your hypothesis if it came out one way, and which would disconfirm your hypothesis if it came out the other way -- but where you can predict the outcome in advance. It's easy to not realize you are doing this. You'll appear to provide significant evidence for your hypothesis, but actually you've cherry-picked your evidence before even looking at it; you knew enough about the world to know where to look to see what you wanted to see.

Suppose Dr. Y studies a rare disease, Swernish syndrome. Many scientists have formed an intuition that Swernish syndrome has something to do with a chemical G-complex. Dr. Y is thinking on this one night, when the intuition crystallizes into G-complex theory, which would provide a complete explanation of how Swernish syndrome develops. G-complex theory makes the novel prediction that G-complex in the bloodstream will spike during early onset of the disease; if this were false, G-complex theory would have to be false. Dr. Y does the experiment, and finds that the spike does occur. No one has measured this before, nor has anyone else put forward a model which makes that prediction. However, it happens that anyone familiar with the details of Dr. Y's experimental results over the past decade would have strongly suspected the same spike to occur, *whether or not* they endorsed G-complex theory. Does the experimental result constitute significant evidence?

This is a subtle kind of double-counting of evidence. You have enough evidence to know the result of the experiment; also, your evidence has caused you to generate a hypothesis. You cannot then claim the success of the experiment as more evidence for your hypothesis: you already know what would happen, so it can't alter the certainty of your hypothesis.

If we're dealing only with personal rationality, we could invoke conservation of expected evidence again: if you already predict the outcome with high probability, you cannot simultaneously derive much evidence from it. However, in group rationality, there are plenty of cases where you want to predict an experiment in advance and then claim it as evidence. *You* may already be convinced, but you need to convince skeptics. So, we can't criticize someone *just* for being able to predict their experimental results in advance. That would be absurd. The problem is, *the hypothesis isn't what did the work of predicting the outcome.* Dr. Y had general world-knowledge which allowed him to select an experiment whose results would be in line with his theory.

To Dr. Y, it just feels like "if I am right, we will see the spike. If I am wrong, we won't see it." From the outside, we might be tempted to say that Dr. Y is not "trying hard enough to falsify G-complex theory". But how can Dr. Y use this advice to avoid the mistake? A hypothesis is an explicit model of the world, which guides your predictions. When asked to *try to falsify*, though, what's your guide? If you find your hypothesis very compelling, you may have difficulty imagining how it could be false. A hypothesis is solid, definite. The negation of a hypothesis includes *anything else*. As a result, "try to falsify your hypothesis" is very vague advice. It doesn't help that the usual practice is to test against a null hypothesis. Dr. Y tests against the spike not being there, and thinks this sufficient.

Implicit Knowledge

Part of the problem here is that it should be very clear what could and could not have been predicted. There's an interaction between your general world knowledge, which is not explicitly articulated, and your scientific knowledge, which is.

If all of your knowledge was explicit scientific knowledge, many biases would disappear. You couldn't possibly have hindsight bias; each hypothesis would predict the observation with a precise probability, which you can calculate.

Similarly, the failure mode I'm describing would become impossible. You could easily notice that it's not really your new hypothesis doing the work of telling you which experimental result to expect; you would know exactly what other world-knowledge you're using to design your experiment.

I think this is part of why it is useful to orient toward gear-like models. If our understanding of a subject is explicit rather than implicit, we can do a lot more to correct our reasoning. However, we'll always have large amounts of implicit, fuzzy knowledge coming in to our reasoning process; so, we have to be able to deal with that.

Is "Sufficient Novelty" The Answer?

In some sense, the problem is that Dr. Y's experimental result isn't novel enough. It might be a "novel prediction" in the sense that it hasn't been explicitly predicted by anyone, but it is a prediction that *could* have been made without Dr. Y's new hypothesis. Extraordinary claims require extraordinary evidence, right? It isn't enough that a hypothesis makes a prediction which is new. The hypothesis should make a prediction which is *really surprising.*

But, this rule wouldn't be any good for practical science. How surprising something is is too subjective, and it is too easy for hindsight bias to make it feel as if the result of the experiment could have been predicted. Besides: if you want science to be able to provide compelling evidence to skeptics, you can't throw out experiments as unscientific just because most people can predict their outcome.

Method of Multiple Hypotheses

So, how could Dr. Y have avoided the mistake?

It is meaningless to confirm or falsify a hypothesis in isolation; all you can really do is provide evidence which helps distinguish *between* hypotheses. This will guide you away from "mundane" tests where you actually could have predicted the outcome without your hypothesis, because there will likely be many other hypotheses which would be able to predict the outcome of that test. It guides you toward corner cases, where otherwise similar hypotheses make very different predictions.

We can unpack "try to falsify" as "come up with as many plausible alternative hypotheses as you can, and look for experiments which would rule out the others." But actually, "come up with alternative hypotheses" is more than an unpacking of "try to falsify"; it shifts you to trying to distinguish between many hypotheses, rather than focusing on "your" hypothesis as central.

The *actual, exactly correct* criteria for an experiment is its value-of-information. "Try to falsify your hypothesis" is a lousy approximation of this, which judges experiments by how likely they are to provide evidence against your hypothesis, or the likelihood ratio against your hypothesis in the case where the experiment doesn't go as your hypothesis predicts, or something. Don't optimize for the wrong metric; things'll tend to go poorly for you.

Some might object that trying-to-falsify is a good heuristic, since value of information is too difficult to compute. I'd say that a much better heuristic is to pretend distinguishing the right hypothesis is equally valuable in all cases, and look for experiments that allow you to maximally differentiate between them. Come up with as many possibilities as you can, and try to differentiate between the most plausible ones.

Given that the data was already very suggestive of a G-complex spike, Dr. Y would most likely generate other hypotheses which also involve a G-complex spike. This would make the experiment which tests for the spike uninteresting, and suggest other more illuminating experiments.

I think "coming up with alternatives" is a somewhat underrated debiasing technique. It is discussed more in Heuer's *Psychology of Intelligence Analysis* and Chamberlin's *Method of Multiple Working Hypotheses*.

Epistemic Spot Check: A Guide To Better Movement (Todd Hargrove)

This is part of an ongoing series assessing where the epistemic bar should be for self-help books.

Introduction

Thesis: increasing your physical capabilities is more often a matter of teaching your neurological system than it is anything to do with your body directly. This includes things that *really really* look like they're about physical constraints, like strength and flexibility. You can treat injuries and pain and improve performance by working on the nervous system alone. More surprising, treating these physical issues will have spillover effects, improving your mental and emotional health. A *Guide To Better Movement* provides both specific exercises for treating those issues and general principles that can be applied to any movement art or therapy.

The first chapter of this book failed spot checking pretty hard. If I hadn't had a very strong recommendation from a friend ("I didn't take pain medication after two shoulder surgeries" strong), I would have tossed it aside. But I'm glad I kept going, because it turned out to be quite valuable (this is what triggered that meta post on epistemic spot checking). In accordance with the previous announcement on epistemic spot checking, I'm presenting the checks of chapter one (which failed, badly), and chapter six (which contains the best explanation of pain psychology I've ever seen), and a review of model quality. I'm very eager for feedback on how this works for people.

Chapter 1: Intro (of the book)

Claim: "Although we might imagine we are lengthening muscle by stretching, it is more likely that increased range of motion is caused by changes in the nervous system's tolerance to stretch, rather than actual length changes in muscles. " (p. 5).

Overstated, weak. (PDF). The paper's claims to apply this up to 8 weeks, no further. Additionally, the paper draws most (all?) of its data from two studies and it doesn't give the sample size of either.

Claim: "Research shows the forces required to deform mature connective tissue are probably impossible to create with hands, elbows or foam rollers." (p. 5).

Misleading. (Abstract). Where by "research" the Hargrove means "mathematical model extrapolated from a single subject".

Claim: "in hockey players, strong adductors are far more protective against groin strain than flexible adductors, which offer no benefit" (p. 14).

Misleading. (Abstract) Sample size is small, and the study was of the relative strength of adductor to abductor, not absolute strength.

Claim: "Flexibility in the muscles of the posterior chain correlates with slower running and poor running economy." (p. 14).

Accurate citation, weak study. (Abstract) Sample size: 8. *Eight*. And it's correlational.

[A number of interesting ideas whose citations are in books and thus inaccessible to me]

Claim: "…most studies looking at measurable differences in posture between individuals find that such differences do not predict differences in chronic pain levels." (p. 31).

Accurate citation. (Abstract). It's a metastudy and I didn't track down any of the 54 studies included, but the results are definitely quoted accurately.

Chapter 6: Pain

Claim: "Neuromatrix" approach to pain means the pattern of brain activity that create pain, and that pain is an output of brain activity, not an input (p93).

True, although the ability to correctly use definitions is not very impressive.

Claim: "If you think a particular stimulus will cause pain, then pain is more likely. Cancer patients will feel more pain if they believe the pain heralds the return of cancer, rather than being a natural part of the healing process." (p93).

Correctly cited, small sample size. (Source 1, source 2, TEDx Talk).

Claim: *Psychological states associated with mood disorders (depression, anxiety, learned helplessness, etc) are associated with pain (p94).*

True, (source), although it doesn't look like the study is trying to establish causality.

Claim: *Many pain-free people have the kinds of injuries doctors blame pain on (p95).*

True, many sources, all with small sample sizes. (source 1, source 2, source 3, source 4, source 5)

Claim: On taking some cure for pain, relief kicks in before the chemical has a chance to do any work (p98)

True. His source for this was a little opaque but I've seen this fact validated many other places.

Claim: we know you can have pain without stimulus because you can have arm pain without an arm (p102).

True, phantom limb pain is well established.

Claim: some people feel a heart attack as arm pain because the nerves are very close to each other and the heart basically never hurts, so the brain "corrects" the signal to originating in the arm (p102).

First part: True. Explanation: unsupported. The explanation certainly makes sense, but he provides no citations and I can't find any other source on it.

Claim: Inflammation lowers the firing threshold of nociceptors (aka sensitization) (p102).

True (source).

Claim: nociception is processed by the dorsal horn in the spine. The dorsal horn can also become sensitized, firing with less stimulus than it otherwise would. Constant activation is one of the things that increases sensitivity, which is one mechanism for chronic pain (p103).

True (source).

Claim: people with chronic pain often have poor "body maps", meaning that their mental model of where they are in space is inaccurate and they have less resolution when assessing where a given sensation is coming from (p107).

Accurate citation (source). This is a combination of literature review and reporting of novel results. The novel results had a sample of five.

Claim: The hidden hand in the rubber hand illusion experiences a drop in temperature (p109).

Accurate citation, tiny sample size (source). This paper, which is cited by the book's citation, contains six experiments with sample sizes of fifteen or less. I am torn between dismissing this because cool results with tiny sample sizes are usually bullshit, and accepting it because it is super cool.

Claim: "a hand that has been disowned through use of the rubber hand illusion will suffer more inflammation in response to a physical insult than a normal hand." (p. 109).

Almost accurate citation (source). The study was about histamine injection, not injury per se. Insult technically covers both, but I would have preferred a more precise phrasing. Also, sample size 34.

Claim: People with chronic back pain have trouble perceiving the outline of their back (p. 109).

Accurate citation, sample size six (pdf).

Claim: "Watching the movements in a mirror makes the movements less painful [for people with lower back pain]." (p. 111). Better Movement. Kindle Edition.

Accurate citation, small sample size (source).

Model Quality

Reminder: the model is that pain and exhaustion are a product of your brain processing a variety of information. The prediction is that improving the quality of processing via the principles explained in the book can reduce pain and increase your physical capabilities.

Simplicity: Good. This is not actually simple model, it requires a ton of explanation to a layman. But most of its assumptions come from neurology as a whole; the leap from "more or less accepted facts about neurology" to this model is quite small.

Explanation Quality: Fantastic. I've done some reading on pain psychology, much of which is consistent with *Guide…,* but *Guide…* has by far the best explanation I've read.

Explicit Predictions: Good, kept from greatness only by the fact that brains and bodies are both very complicated and there's only so much even a very good model can do.

Useful Predictions: Okay. The testable prediction for the home-reader is that following the exercises in the back of the book, or going to a Feldenkrais class, will treat chronic pain, and increase flexibility and strength. Since the book itself admits that a lot of things offer short term relief but don't address the real problem, helping immediately doesn't prove very much.

Acknowledging Limitations: Poor. *GTBM* doesn't have the grandiose vision of some cure-all books, and repeatedly reminds you that your brain being involved doesn't mean your brain is in control. But there's no sentence along the lines of "if this doesn't work there's a mechanical problem and you should see a doctor."

Measurability: poor. This book expects you to put in a lot of time before seeing results, and does not make a specific prediction of the form they will come in. Worse, I don't think you can skip straight to the exercises. If I hadn't read the entire preceding book I wouldn't have approached them in the correct spirit of attention and curiosity.

Hmmm, if I'd assigned a gestalt rating it would have been higher than what I now think is merited based on the subscores. I deliberately wrote this mostly before trying the exercises, so I can't give an effectiveness score. If you do decide to try it, please let me know how it goes so I can further calibrate my reviews to actual effectiveness.

You might like this book if…

…you suffer from chronic pain or musculoskeletal issues, or find the mind-body connection fascinating.

This post supported by Patreon.