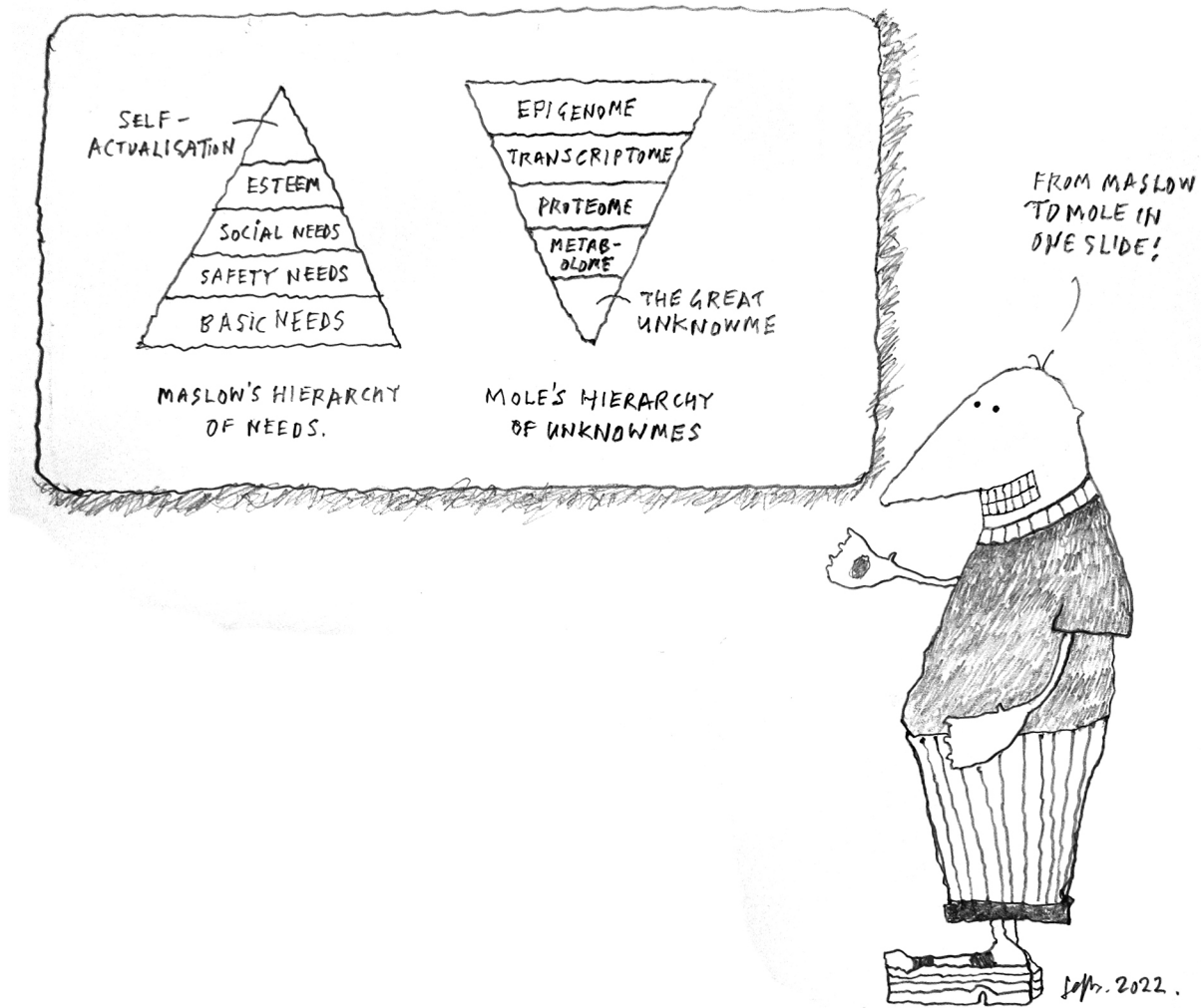


STICKY WICKET

The unknown

Mole



Original artwork by Pete Jeffs - www.peterjeffsart.com

Where in the world did I put my phone? I was reading a paper, and then I had it in my hand when I went to the kitchen to get a pretzel (yes, I like pretzels; the ones we have here are hard and crunchy, but I also love the pretzels that I have had in a tent at the Oktoberfest. That's great fun; millions of people listening to Bavarian bands while drinking enormous quantities of beer, the latter being the whole point. If we had anything like this in my country, it would degenerate into violent chaos, but in Bavaria, it all seems to work. Each year, I join a 'think tank' of scientists at Oktoberfest to think deep thoughts while drinking beer; the pandemic put this on hold for a couple of years, but I am going *this* year, which will be great fun. We should probably call our group a 'drink tank'. What was I doing? Oh yes, looking for my phone). It isn't there, and it isn't in the bedroom. Or on the counter

by the sink. Oh! Here it is, in my other hand. Now, who was I trying to call?

There is a maxim that you always find lost items in the last place you look. And it has frequently been pointed out (most recently in a fortune cookie I received after a nice meal in a Chinese restaurant) that this is because you stop looking after you find it. A satisfying hypothesis, and while I don't know the extent to which it has been tested, it might be true. Sometimes, we don't need experimental testing to know something.

So why am I talking about fortune cookies and pretzels (maybe I'm hungry)? Well, it is a beautiful day, so I have been doing what I often do, sitting outside reading papers. And in the past few months, I have noticed a trend in the first few sentences (often the abstract) of many papers I peruse. Probably it has been around much longer,

but lately it has been bugging me. So, I want to talk about it, assuming (as I do) that it's okay with you.

It starts with a statement of something that the reader will agree is important. Then the next sentence mentions something, a gene, a protein, a metabolite, or whatever, and then notes that the role of that something in the first something is unknown. For example, (I'm making this up) "Sleeping accounts for ~33% of our daily life. Many of us use cell phones during much of the waking day, but the role of cell phones during sleep is unknown." (Oh, right, *that's* why I was looking for my phone). Sleep is important, phones are important (to many of us), but that doesn't mean that the relationships between sleep and phones is important. It's just unknown. (Now substitute your favorite disease, cell type, molecular process, or pretzel – okay, not pretzel – for 'sleep' and your favorite molecule, gene, cell type, or snack – sorry, not snack – for 'cell phone'. See if this is a good start for a paper).

To do this biomedical research thing that we do, many of us have to explore the unknown. And sometimes the 'this is important and that is important, but the role of that in this is unknown' approach is indeed a route to making the unknown known an interesting proposition. If a drug treatment was found to effectively treat six related diseases, but we don't know if it would help in a seventh related disease, then by all means it makes sense to ask the question. However, if a molecule plays roles in six cellular pathways and you decide to find out if it functions in a seventh one, you could do that, but it isn't very likely that you'll convince me that doing so is important, *a priori*.

In the majority of the studies that I've read that include the 'role of that in this is unknown' introduction, I don't think they started out like that. What I suspect is that the researchers approached a problem and then used an omics sort of approach to identify things that might help to address the problem, and then found out that, yes, something they found seemed to shed some light on what they were studying. Or maybe it was a lucky guess. But when it came time to write the paper, a *reason* to look at that thing wasn't obvious, so they simply stated that 'the role of Ptpx43a in disease of the left eyelid is unknown'. (Yes, I made that up. We all know the role of Ptpx43a in disease of the left eyelid. It's nothing).

Once upon a long time ago, we used to do research by framing and then testing a hypothesis. We would read a lot of papers and think deeply about a problem, and then think of a possible solution, and then design experiments to see if we were right (or, depending on which philosopher of science you prefer, see if we were wrong). Then, if it appeared that we were on the right track, we would design more experiments to follow the story to see where it went. Some of us still do that (complete with the 'design experiments to see if we are right/wrong' part). When it came time to write the paper, we explained how we arrived at our hypothesis and what we did to test

it. Easy. Since the time of Francis Bacon and the scientific method, this was how science was done.

That was then. Now, we have incredible tools to explore the unknown. We can dissect model systems and patient samples to interrogate the genome, epigenome, and transcriptome at the single-cell level, and (coming soon!) the single-cell proteome, metabolome, and any other 'ome' you can think of. And we are learning to do this spatially, 'seeing' what is going on in each cell in a healthy or diseased tissue. It is all so complicated that we need expert data scientists, and often they have to resort to 'machine learning'. (And when a machine 'learns' to analyze data, it generally can't tell us how the results were actually obtained). And through all of this, we sometimes find out that a 'thing' (gene, protein, metabolite, snack food – no, not a snack food) went up or down and we speculate that 'hey, maybe that thing is important'. And since the 'role' of that thing in our system was unknown, it seems perfectly reasonable to point this out in the introduction of our paper. Because we don't really know *why* we decided to look at it in the first place.

But here is the point (there's a point? C'mon, you know there's a point. This isn't our first rodeo). We like to say that 'research need not be hypothesis driven', indeed, our funding agencies explicitly tell reviewers this. But there *is* a hypothesis in every one of these studies. The hypothesis, generally put, is this: I hypothesize that by accumulating very large amounts of data and figuring out ways to analyze it, I will gain fundamental insights into the problem under consideration. We test it by doing this accumulation and analysis and see where it gets us. And sometimes it gets us pretty far and we discover new things. But in every case, *every* case (I can think of) we end up with a new, and rather informed hypothesis. And in those cases where we go on to *test* the new hypothesis, we find out if we (maybe) are on the right track. We have new tools to explore the unknown, but we shouldn't confuse those tools with something other than the scientific method. Listen to your inner Francis.

So please, the next time you are tempted to write 'the role of that in this is unknown' in the abstract or introduction of your paper, think of your reader. Why should anyone actually *care*? Not everything that is unknown needs to be known. Science is hard and it is expensive (and omics can be *very* expensive). But readers' time is also expensive. Don't be lazy; take the time to explain why this is a burning question. And if you don't know why you asked the question in the first place (for example, someone told me to do it) find out a better reason than 'it was unknown').

In 1924, George Mallory was asked why he wanted to climb Mt Everest, and he replied, "because it's there". (This quote is often attributed to Edmond Hillary, who did not say it. George did.) He was sadly lost, never heard from again. The same may be said of this paper I was just reading. But then again, I don't know.

I guess it's unknown. Now, what was I looking for again?