

The Physicist turned Data Scientist

Behavioral Analysis - three strengths and their dark sides

Takes one to know one!

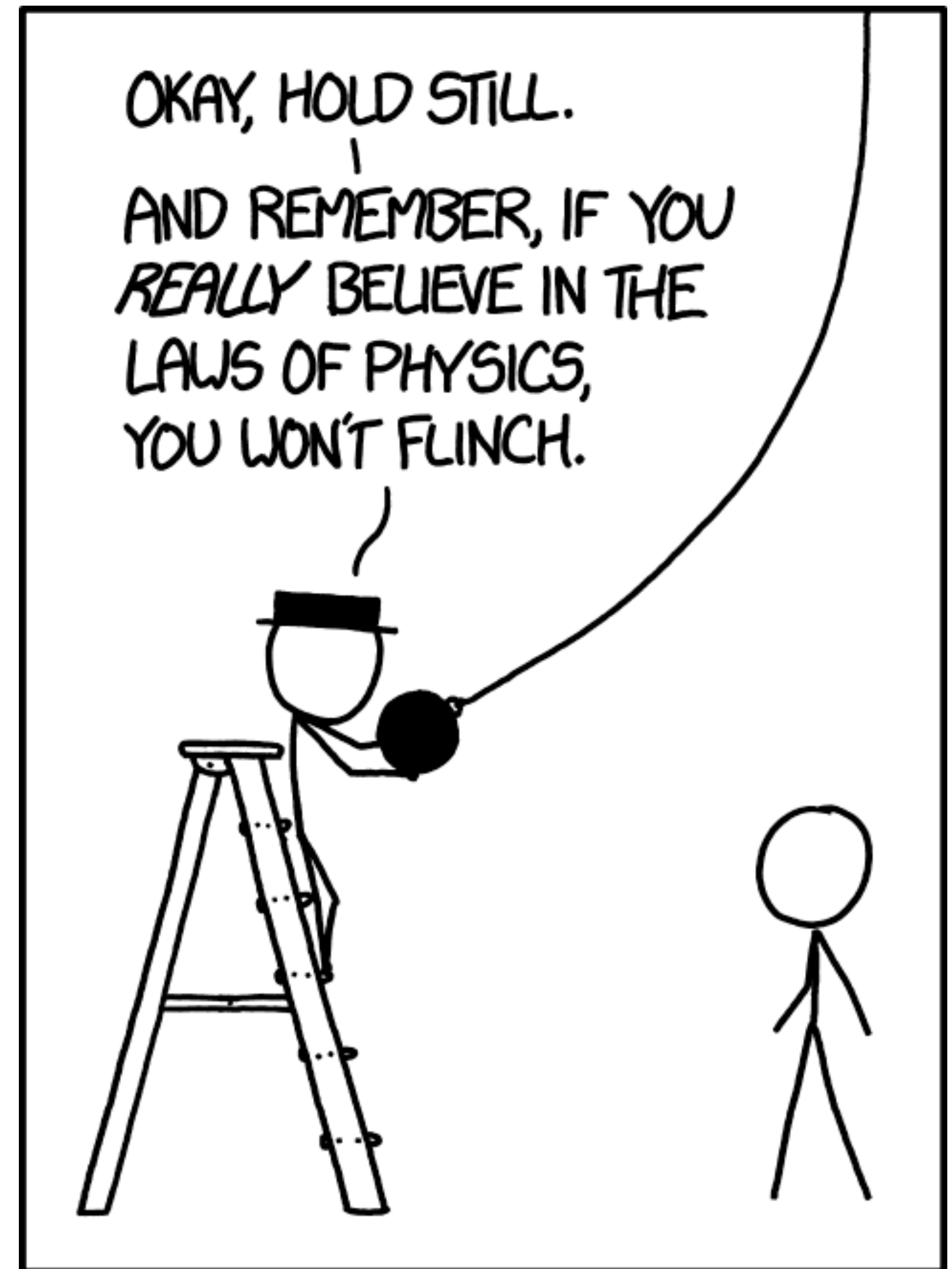
- Particle physics, CMS collaboration at the LHC
- PhD 2012, postdoc until 2018, DS ever since



BlueCross® BlueShield®
Illinois • Montana • New Mexico
Oklahoma • Texas

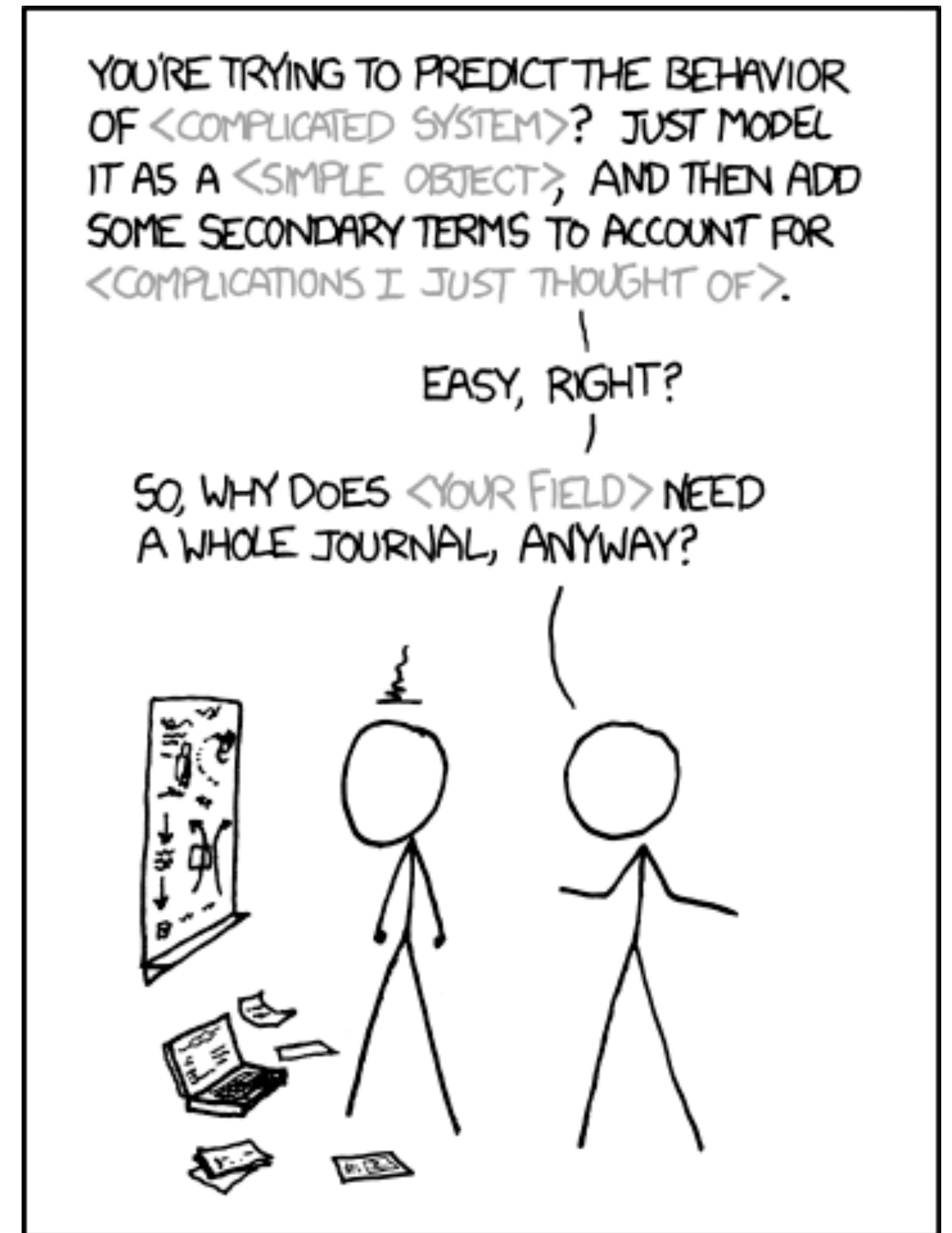
Strength #1 - Confidence

- Confidence in science
- Confidence in their own abilities
- Not scared of math



Dark side #1 - Hubris

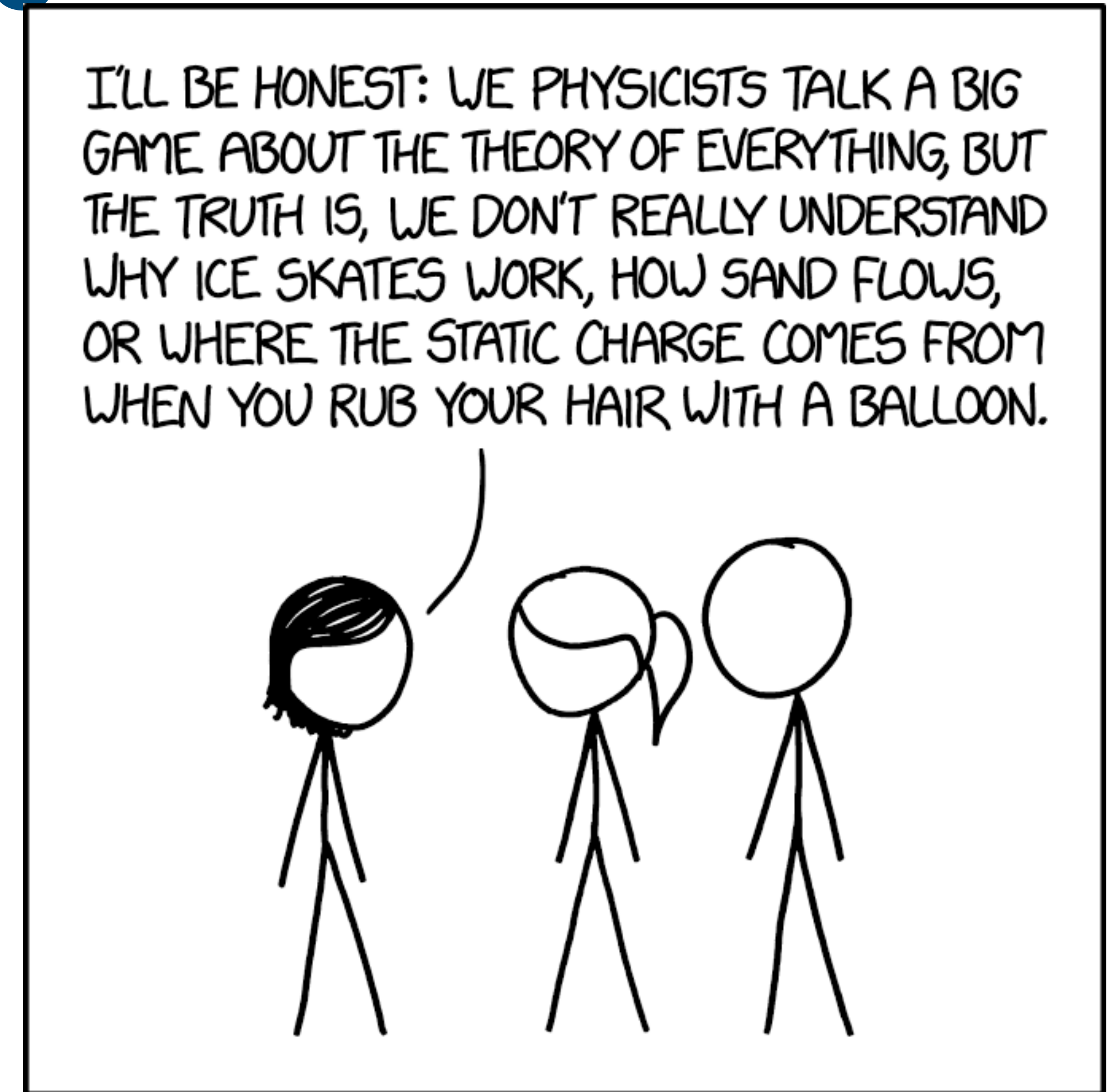
- Look down on “less pure” fields
- Lack of appreciation for non-technical skills
- Idea that other job functions exist solely to support them



LIBERAL-ARTS MAJORS MAY BE ANNOYING SOMETIMES, BUT THERE'S *NOTHING* MORE OBNOXIOUS THAN A PHYSICIST FIRST ENCOUNTERING A NEW SUBJECT.

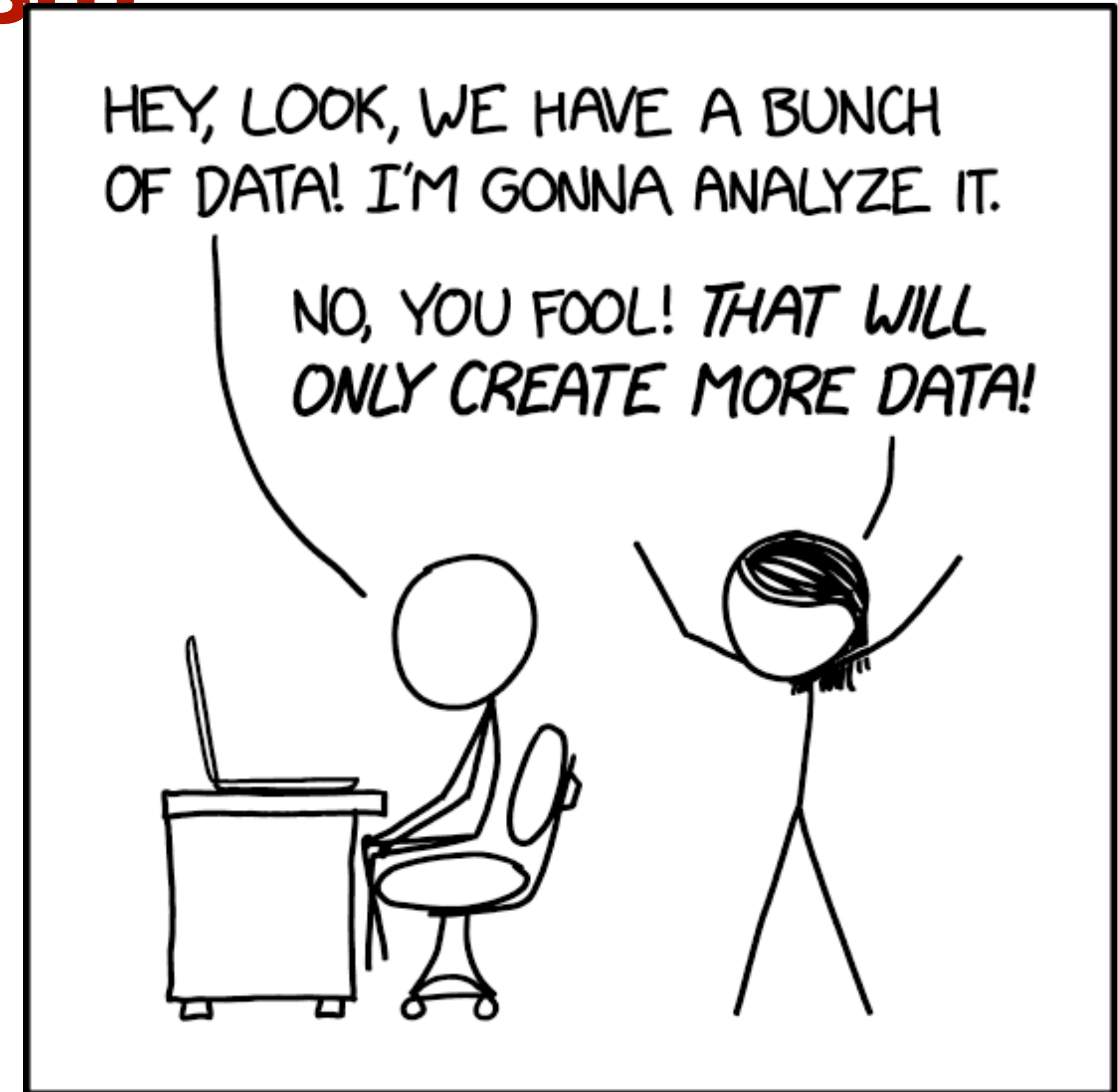
Strength #2 - Perseverance

- Understanding that the answers may be murky
- Comfort with tedious tasks requiring attention to detail
 - Think pages upon pages of algebra
- Healthy skepticism of easy answers



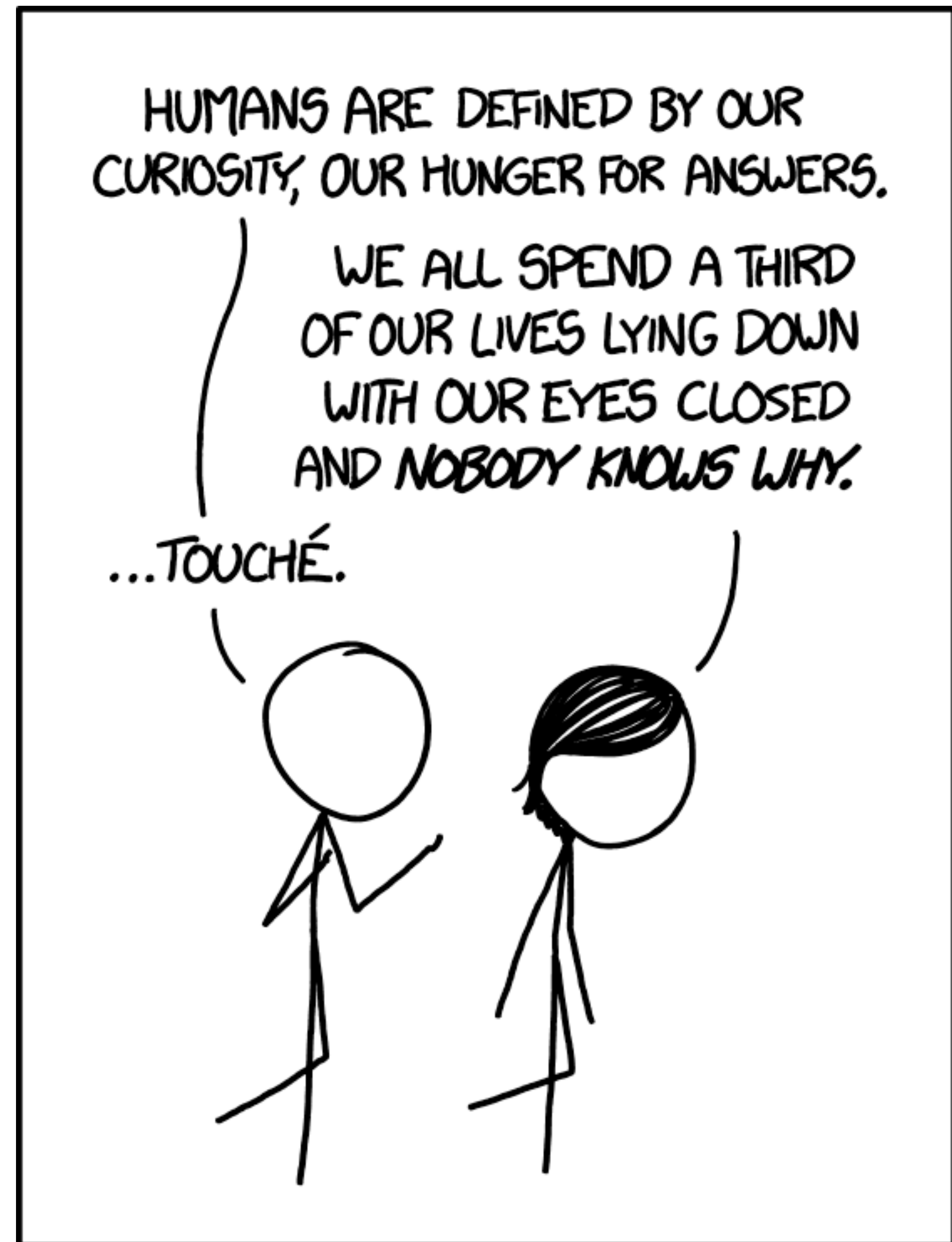
Dark side #2 - Perfectionism

- Too much *depth*
- Analysis paralysis
- Letting the perfect be the enemy of the good



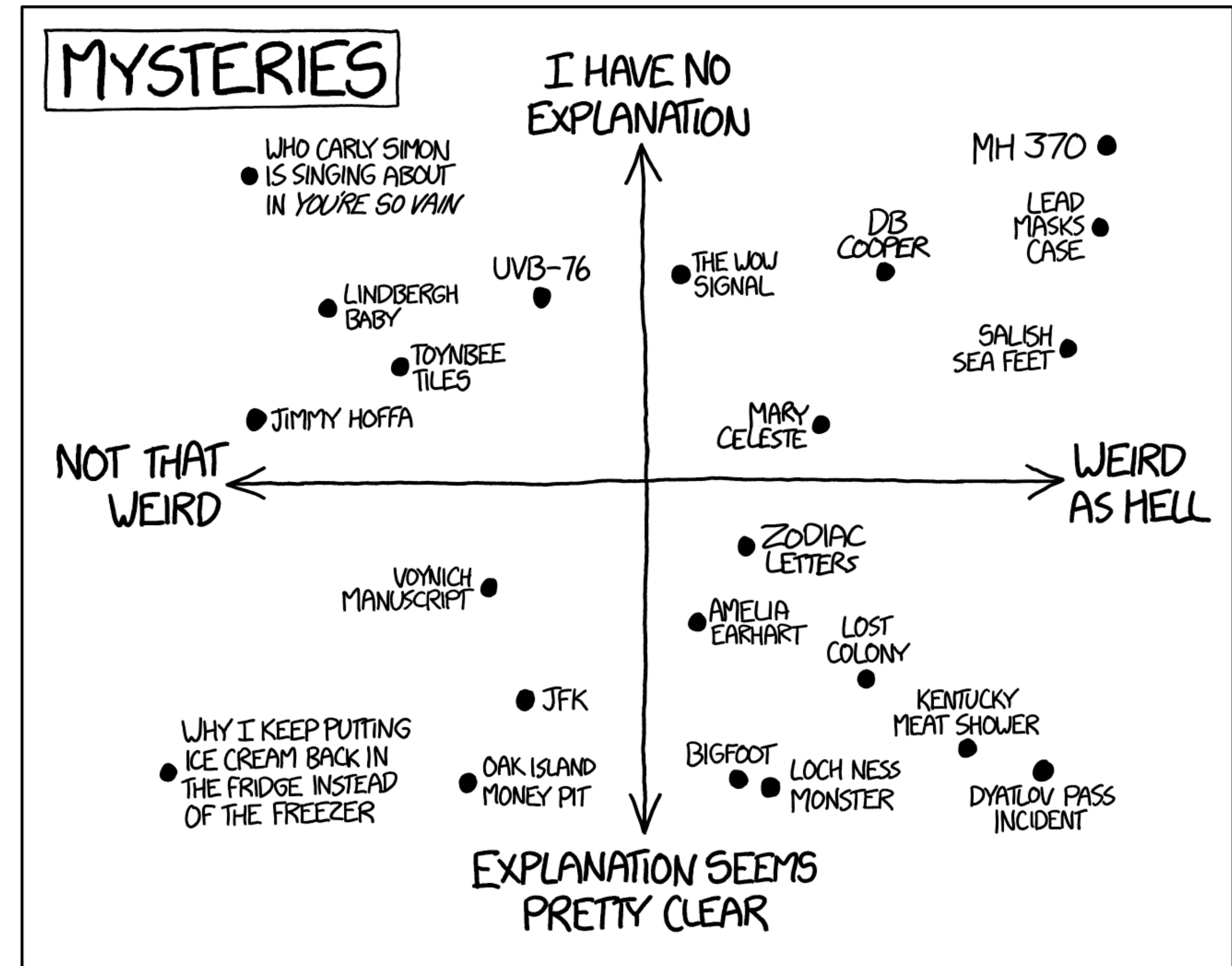
Strength #3 - Curiosity

- Desire to build *intuition*
- Interest in input data, not just fancy modeling
- Can apply the scientific method, generate and test hypotheses



Dark side #3 - Distractibility

- Too much *breadth*
- Focus on accumulating knowledge for its own sake
- Easy to lose sight of the *value* a DS project is creating



Bonus Slide: If you really want some tools to learn

- **Python for data analysis**
 - `pandas`, `matplotlib`, `sklearn`
- **SQL**
 - Can learn a lot of the relevant concepts in `pandas` as well
- **Binary classification models with tree-based algorithms**
 - Accounts for ~90% of what we do in DS @ BCBS
- **Git/github**
 - Generally - writing code with the assumption others will read it, use it, and develop on top of it
- **Get out there and build something!**