Welcome to Unifying Data Science *(Continued)*

Nick Eubank

Part Two: Unifying Data Science

How did Data Science become a thing?

- Academic research is organized into silos:
  - Computer Science
  - Statistics
  - Economics
  - Political Science
  - Engineering

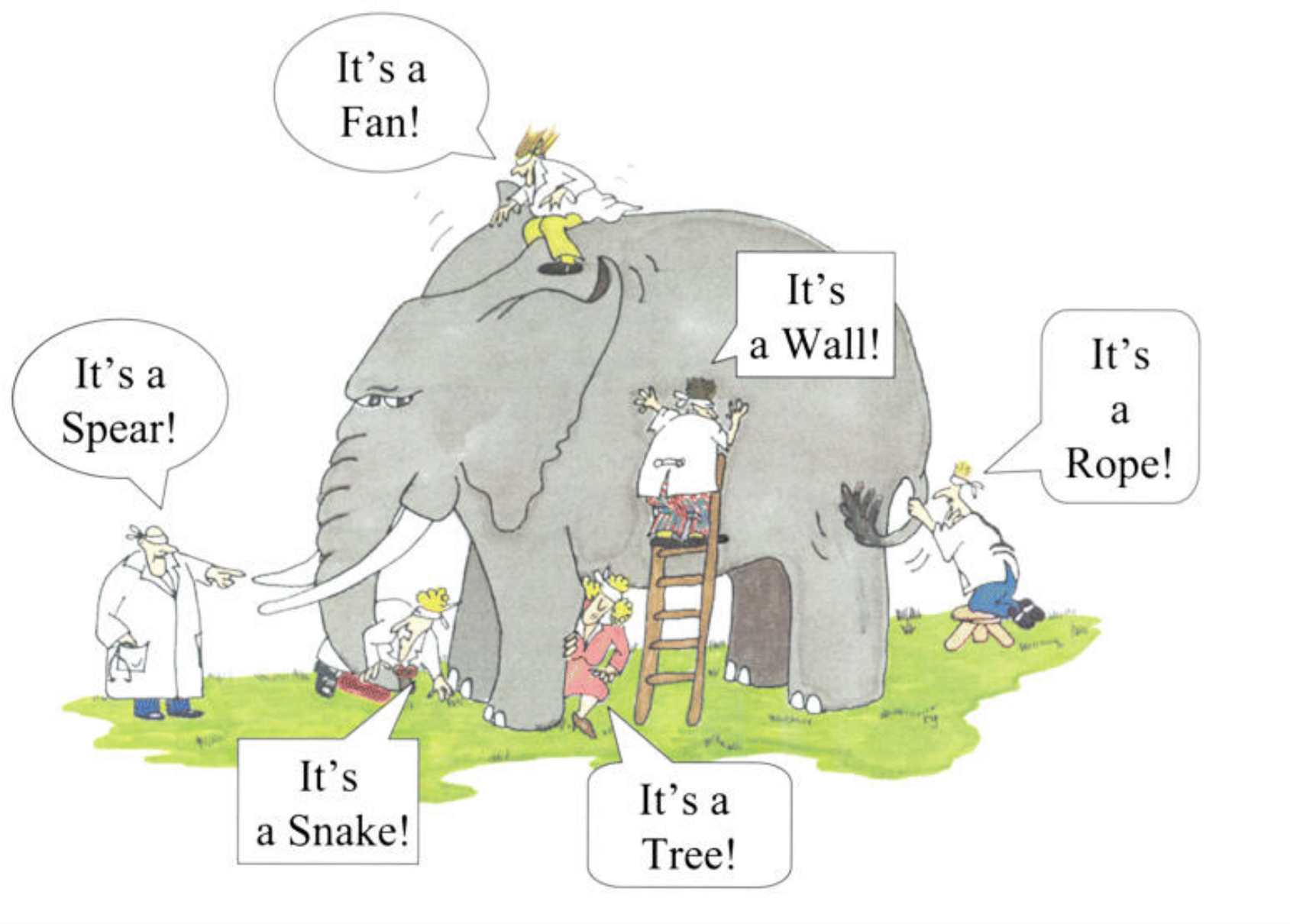
$\Rightarrow$ Development of new tools occurred *within* each silo.

Where are we today?

Very little cross-pollination across silos

- Lots of duplication of development.
- Every silo has its own vocabulary.
- Each silo has focused on the aspects most relevant to their applications. e.g.:
  - CS likes to classify things and make predictions, don't care how model works
  - Social scientists like to make causal statements, don't care about predictive power

Blind Men and the Elephant



⇒ This is where data science is *now*.

What do I think Data Science should be?

An effort to unify the development of quantitative methods

$\rightarrow$ Recognize the elephant

This Class

Discipline of learning how best to answer questions using quantitative data.

This Class

1. Introduce a taxonomy of questions
   Descriptive, causal, predictive
2. For each class of questions, we will discuss:
   - Intrinsic challenges to answering each class of questions
   - What tools are best suited to each type of question

By the end of the course, you should know when to reach for...

- Unsupervised machine learning
- Supervised machine learning
- Range of causal inference techniques
- Other approaches to descriptive analysis

This Class

The tool you use should be dictated by the question you seek to answer

Part Three: Your Data Science Project

Data Science Project

Over semester, you will also develop a data science project from start-to-finish

- Teams of 3-4,
- On topic of your own choosing.
- Only rule: it has to be causal.

$\rightarrow$ Nice portfolio piece
$\rightarrow$ MIDS first-years: Capstone with training wheels

Data Science Project

Introducing in stages:

- Stakeholder management
- Backwards Design
- Workflow Management
- Presenting to Different Audiences
- Giving Feedback

Who Are We?

I am a empirical / computational social scientist

- PhD in Political Economy, Masters in Economics, BA in Economics and International Relations
- Research on criminal justice, policing, social networks, election administration, gerrymandering, and (in days gone by) international development.

(But I have a pretty strong CS background for a social scientist.)

Nathan Warren & Becky Chen (TA)

- MIDS Second Year Students
- *Extremely* good at this
- Causal inference is a discipline that people spend their careers studying, so they are terrific resources, but also be aware you may hit questions they redirect to me.

Things to Know

- Course site: `http://www.unifyingdatascience.org`
  Contents subject to change!
- Readings are *incredibly* important.
- Reading reflections for every reading.
  Due **Night before class.**
  Reading quizzes are likely to be a regular feature of the class.
- If you don't know git or github, you'll want to learn that early.
  - Data Camp and Practical Data Science tools will be made available
  - Workshops hosted by Library
- Books:
  - *Mastering 'Metrics*
  - *Mostly Harmless Econometrics*

If you have issues...

- With the course material,
- With the course design,
- With learning online,
- With the isolation associated with COVID-life,
- *Or anything else...*

Talk to me!

Phew. That's it!
Questions?