“Big Data” Need(s?) Bigger Theory and Methods

Trevor Self, Ph.D.
International Monetary Fund (IMF) – April 12, 2014

Caveat

- The opinions expressed in this presentation are my own and do not necessarily represent the positions of the International Monetary Fund (IMF) or its member countries.

What are (is?) “Big Data”?

- “Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...” (Dan Ariely, 1/26/2013)
- Digital data with high levels of:
  - Volume
  - Velocity
  - Variety/ Variability

Concerns

- Implications for inclusion and legitimacy?
- The reciprocal relation between technology and informational processing isn't new.
- As always, decisions are improved by informational efficiency, not volume. (quality ÷ quantity)
- Explanation usually matters.

Concerns

- There’s no such thing as “raw data”.
- The data are not the domain, and fetishizing data is dangerous.
- Technology does not absolve us our responsibilities.
- What does “data science” add?
- Over-interpretation

Curricular Recommendations

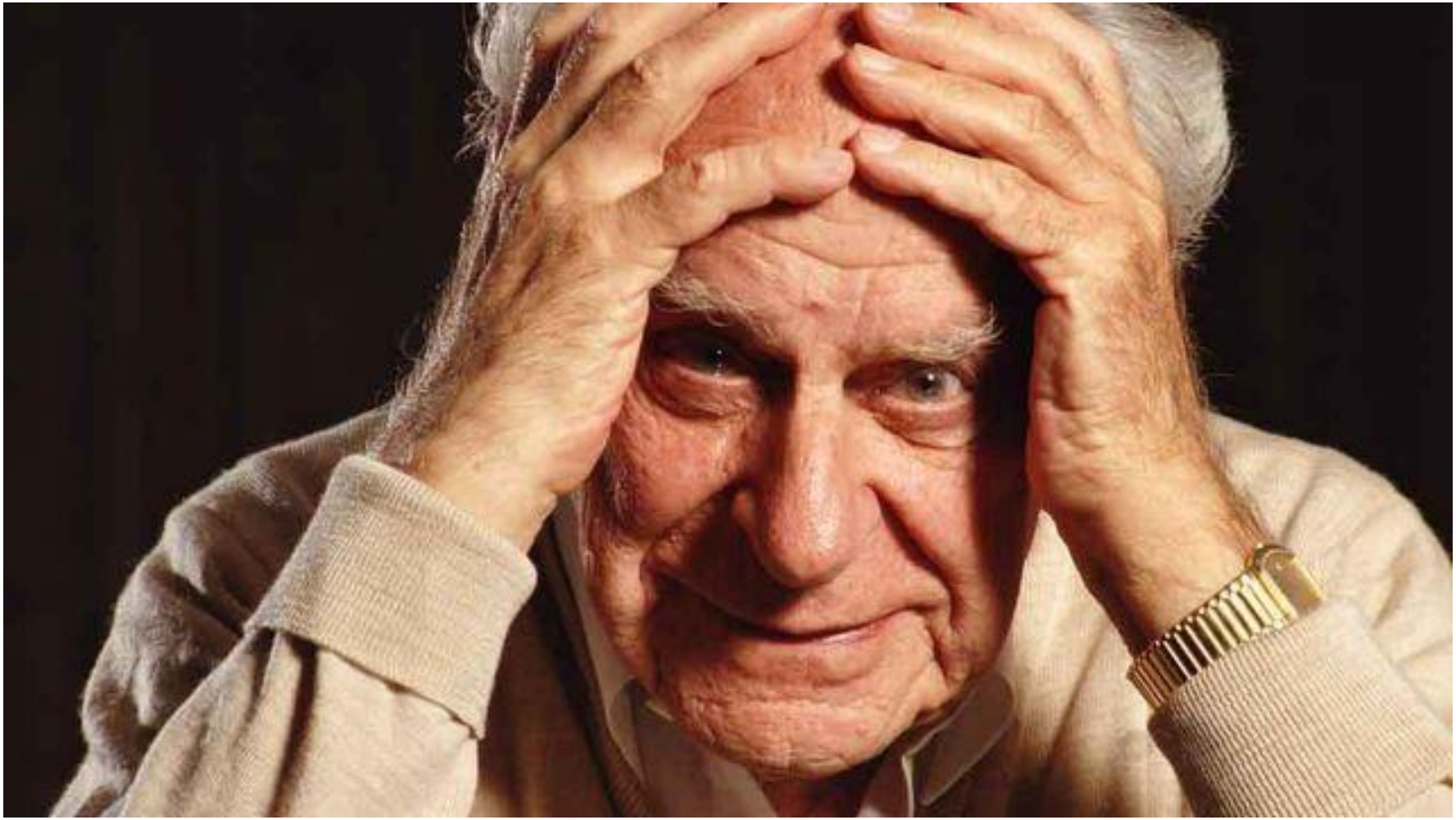
- Good Governance
- Interpretation > Methods > Statistics
- 3 Types of Policy Data
  - Exploratory & Serendipitous (hypothesis formation)
  - Disconfirmatory (scientific)
  - Administrative dashboards
- Threats to validity; cross-validation; ethics
- Meta-analyses
- Multi-trait, multi-method analyses; measurement theory; structural equation modeling

Taking out the trash

- "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete" (Chris Anderson, Editor in Chief, *Wired*, 6/23/2008)
  - "The Petabyte Age is different because more is different. […] With enough data, **the numbers speak for themselves**."
  - "Petabytes allow us to say: 'Correlation is enough.' We can stop looking for models. **We can analyze the data without hypotheses** about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and **let statistical algorithms find patterns where science cannot**."

Taking out the trash

- “I believe that **math is trumping science**. What I mean by that is you don’t really have to know why, you just have to know that if a and b happen, c will happen.” (Vivek Ranadivé, founder & CEO of TIBCO, Techonomy, Nov 13, 2011)
- “Having more data beats out having better models […]” (Edd Dumbill, O’Reilly, 2012)



Sir Karl Popper reacting to the epistemic claims of some “Big Data” proponents.

“With great (1 - β) comes great responsibility.”