Department of Methodology

Department of Methodology Inaugural Lecture

The Challenge of Big Data for the Social Sciences

Professor Kenneth Benoit
Professor of Quantitative Social Research Methods, LSE

Kenneth Cukier
Data Editor,
The Economist

Professor Simon Hix
Chair, LSE

Suggested hashtag for Twitter users: **#LSEdata**

LSE events

CPD CERTIFIED
The CPD Certification Service

Big Data
rerum cognoscere causas

In response to Ken Benoit's Department of Methodology Inaugural Lecture on The Challenge of Big Data for the Social Sciences

Kenneth Cukier

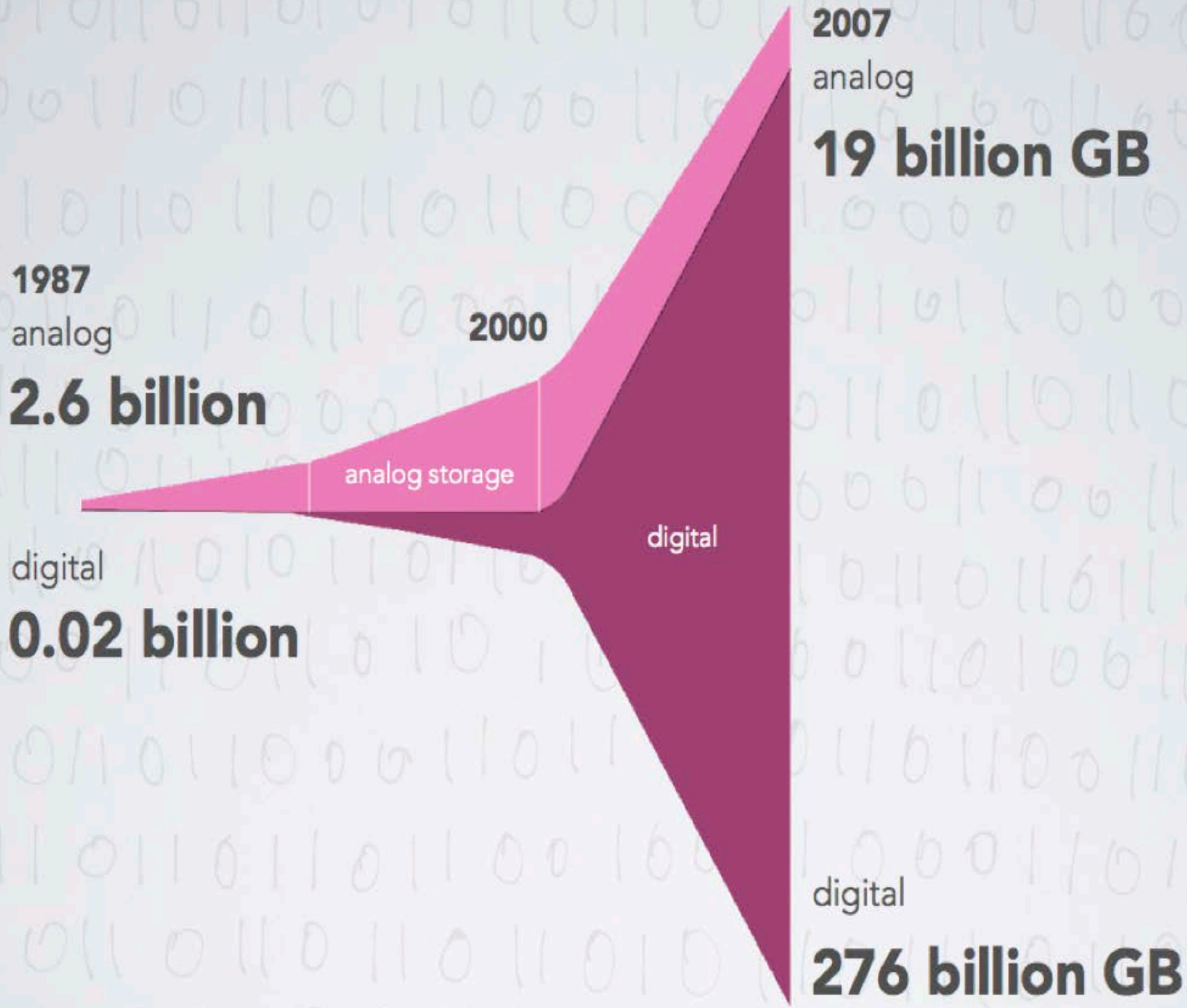
Data Editor, The Economist

Co-author "Big Data: A revolution that will transform how we live, work and think"

London School of Economics and Political Science, February 15, 2015

BIG

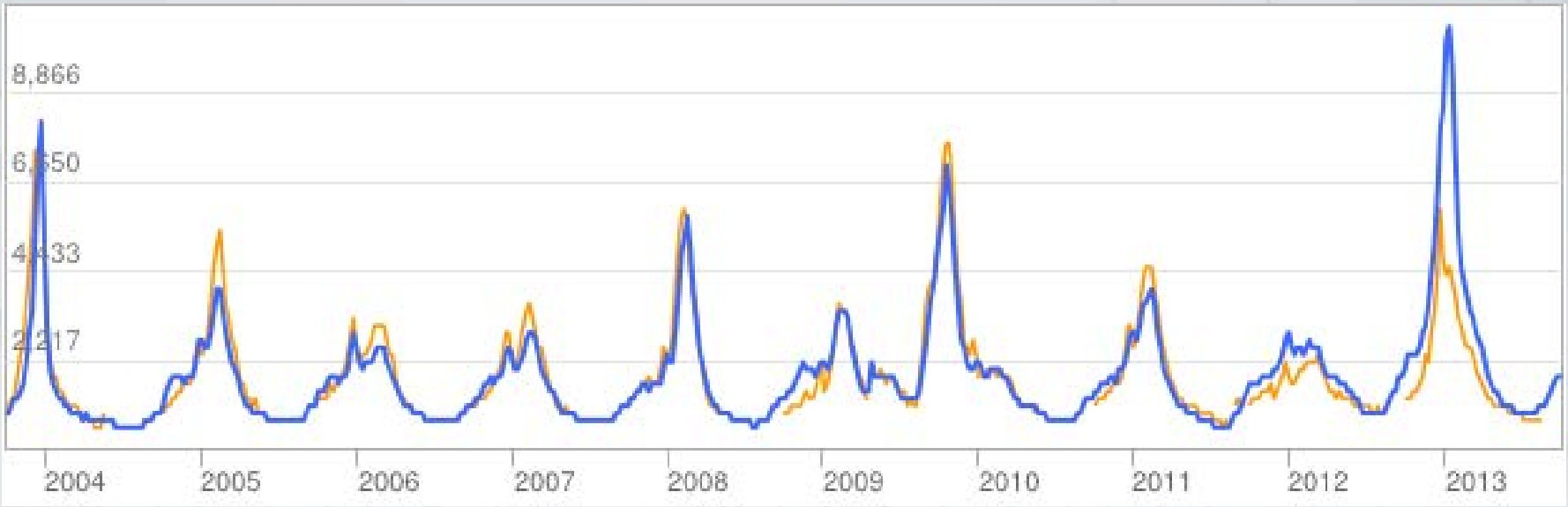
DATA?



Source: "The World's Technological Capacity to Store, Communicate, and Compute Information," Martin Hilbert, Priscila López, Science, 1 April 2011: Vol. 332 no. 6025 pp. 60-65

predict | explain

gft


8,866
2,217
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

h terms that match the propen-
u but are structurally unrelated,
t predict the future, were quite
evelopers, in fact, report weed-
nal search terms unrelated to the
gly correlated to the CDC data,
e regarding high school basket-
is should have been a warning
data were overfitting the small
ses, a standard concern in data
s ad hoc method of throwing
search terms failed when GFT
missed the nonseasonal 2009
–H1N1 pandemic (*2*, *14*). In
ial version of GFT was part flu
winter detector. GFT engineers

Considering the large number of approaches that provide inference on influenza activity (*16*–*19*), does this mean that the current version of GFT is not useful? No, greater value can be obtained by combining GFT with other near–real time health data (*2*, *20*). For example, by combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, we can substantially improve on the performance of GFT or the CDC alone (see the chart). This is no substitute for ongoing evaluation and improvement, but, by incorporating this information, GFT could have largely healed itself and would have likely remained out of the headlines.

Source: D. Lazer, R. Kennedy, G. King, A. Vespignani. “The Parable of Google Flu: Traps in Big Data Analysis,” Science, vol. 343, March 14, 2014

n=all

Schizophrenia GWAS: Number of significant loci

Source: Manolis Kellis, “Importance of Access to Large Populations,” Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice, Cambridge, MA, March 3, 2014

knowledge

ethics, law and technology

Samsung Smart TV privacy policy

Recognition features to you. In addition, Samsung may collect and your device may capture voice commands and associated texts so that we can provide you with Voice Recognition features and evaluate and improve the features. Please be aware that if your spoken words include personal or other sensitive information, that information will be among the data captured and transmitted to a third party through your use of Voice Recognition.

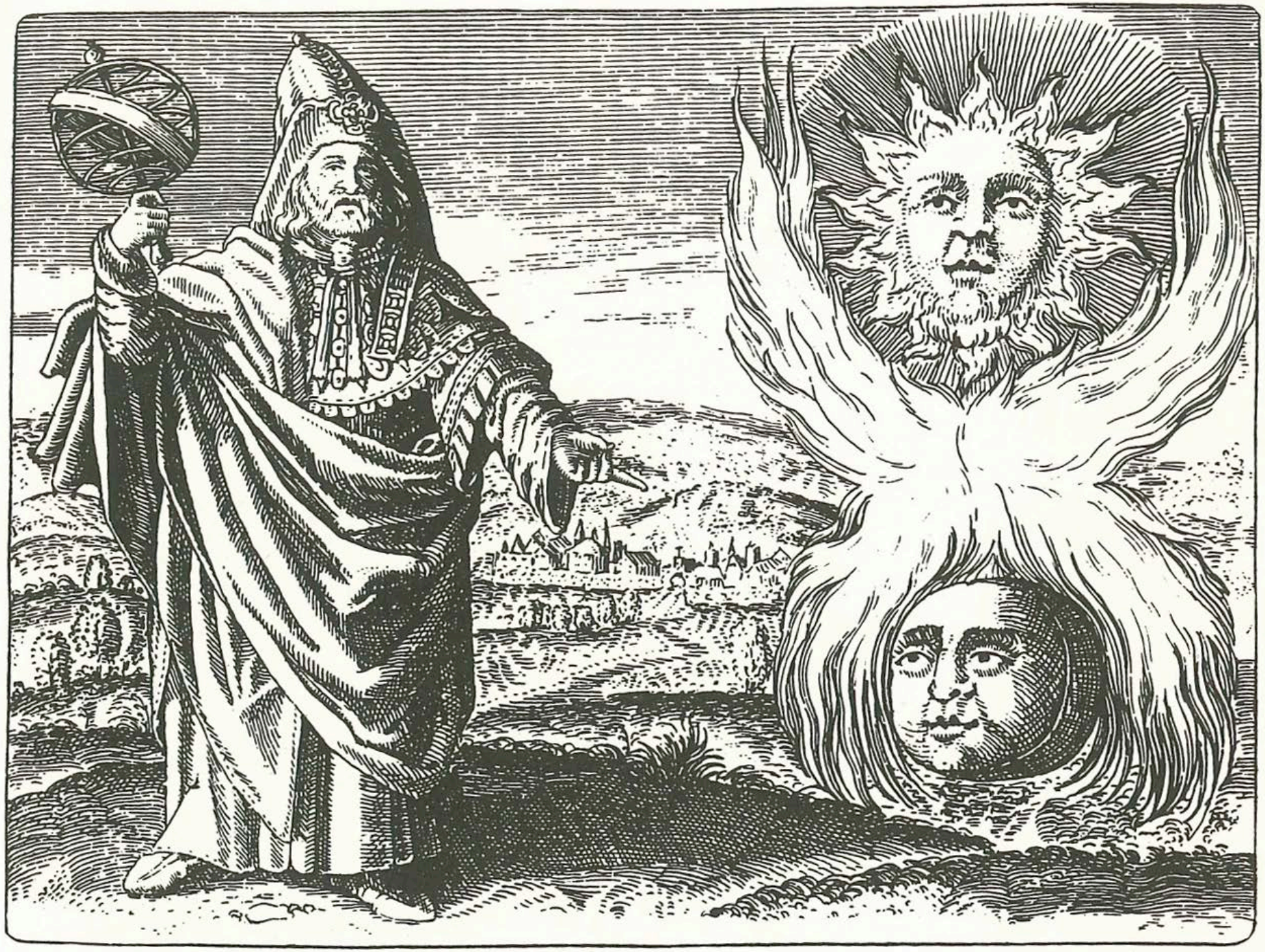
If you do not enable Voice Recognition, you will not be able to use interactive voice recognition features, although you may be able to control your TV using certain predefined voice commands. While Samsung will not collect your spoken word, Samsung may still collect associated texts and other usage data so that

George Orwell, 1984

Behind Winston's back the voice from the telescreen was still babbling away about pig-iron and the overfulfilment of the Ninth Three-Year Plan. The telescreen received and transmitted simultaneously. Any sound that Winston made, above the level of a very low whisper, would be picked up by it, moreover, so long as he remained within the field of vision which the metal plaque commanded, he could be seen as well as heard. There was of course no way of knowing whether you were being watched at any given moment. How often, or on what system, the Thought Police plugged in on any individual wire was guesswork. It was even conceivable that they watched everybody all the time. But at any rate they could plug in your wire whenever they wanted to. You had to live--did live, from habit that became instinct--in the assumption that every sound you made was overheard, and, except in darkness, every movement scrutinized.

change

people & organizations



thank you

KennethCukier@Economist.com

@kncukier

Department of Methodology

Department of Methodology Inaugural Lecture

The Challenge of Big Data for the Social Sciences

Professor Kenneth Benoit
Professor of Quantitative Social Research Methods, LSE

Kenneth Cukier
Data Editor,
The Economist

Professor Simon Hix
Chair, LSE

Suggested hashtag for Twitter users: **#LSEdata**

LSE events

CPD CERTIFIED
The CPD Certification Service