More thinking about less data: A perspective from the 2nd Provence Summer Workshop

M Valenzuela
University of Sydney

D Bartres-Faz
University of Barcelona

E Bullmore
University of Cambridge

A Fjell
University of Oslo

M Maletic-Savetic
Baylor College Of Medicine

See next page for additional authors

Publication Details
More thinking about less data: A perspective from the 2nd Provence Summer Workshop

Abstract
Doppler intuited that a sound’s pitch could be altered by the relative velocity between the source and an observer-70 years later Hubble used the same principle and 42 data points to prove the universe was indeed expanding. Arguably, no other data set of 0.042 Kb has done more to change our understanding of the cosmos. Although modest in volume, it took Hubble several years to acquire these precious numbers.

Keywords
provence, 2nd, perspective, workshop, data, summer, less, about, thinking, more

Disciplines
Education | Social and Behavioral Sciences

Publication Details

Authors
M Valenzuela, D Bartres-Faz, E Bullmore, A Fjell, M Maletic-Savetic, R Martins, N Solowij, and M Yucel

This journal article is available at Research Online: http://ro.uow.edu.au/sspapers/540
More thinking about less data: A perspective from the 2nd Provence Summer Workshop.

Michael Valenzuela1*, David Bartres-Faz2, Ed Bullmore3, Anders Fjell4, Mirjana Maletic-Savetic5, Ralph Martins6, Nadia Solowij7, Murat Yücel8.

Affiliations
1. Regenerative Neuroscience Group, Brain and Mind Research Institute, University of Sydney, Australia
2. University of Barcelona, Spain
3. Cambridge University, United Kingdom
4. University of Oslo, Norway
5. Baylor College of Medicine, United States
6. Edith Cowan University, Australia
7. University of Wollongong, Australia
8. University of Melbourne, Australia

Corresponding Author
Associate Professor Michael Valenzuela
Regenerative Neuroscience Group
Brain and Mind Research Institute
University of Sydney
100 Mallett St Camperdown
NSW 2050
Australia
Doppler intuited that a sound's pitch could be altered by the relative velocity between
the source and an observer – 70 years later Hubble used the same principle and 42 data
points to prove the universe was indeed expanding. Arguably, no other dataset of
0.042Kb has done more to change our understanding of the cosmos. Whilst modest in
volume, it took Hubble several years to acquire these precious numbers.

Nowadays we conduct neuroscience in a state of instant data overload. In a matter of
hours we could produce a structural image of an individual's brain comprising a matrix
of \(256 \times 256 \times 128 = 8,388,608\) data points, a resting-state functional MRI timeseries
(83MB), and from a simple blood sample, derive the person’s genetic sequence by GWAS
(30K), state of gene expression by microarray (another 30K) and metabolomic profile
using any of a number of commercially available chips (1K). We needn't necessarily stop
there. In principle, the number of brain-gene-omic interactions on permutation alone
approaches \(10^{19}\).

It's safe to predict that the ordinary person is not well placed to think about this **uber**
matrix. Recall the human brain 'only' contains \(10^{12}\) synapses. And even if we could
literally combine brain power, it would take a cluster of 10,000,000 networked humans
to allocate just a single synapse to any given interaction. Let's also ignore the fact that
Matlab maxes out on a matrix greater than 8TB, or the pesky issue of Type I error.

Data reduction is therefore one of the key problems for a field trying to better
understand the complexities of the human brain and how and why it sometimes breaks
down. At the Workshop, we discussed two general counter-strategies. The first was
data-driven. For example, computationally intense graph theoretical analyses fit within
a wider systems biology framework that aims to reduce highly multi-dimensional data
into less dimensional more tractable data. The second was having more time to think a
great idea.

Graph theory searches for invariant structures within large matrices, patterns of
covariance that we now appreciate are related to the ability of complex systems to
transfer information. Ed Bullmore identified four areas where graph theory may help
advance neuroscience. Firstly, it provides novel metrics for looking at pre-existing data
in a new way. Secondly, it helps compare and converge results across multiple image
(and non-image) modalities. Thirdly, it may provide a tool for understanding how
mechanistic changes at one level of enquiry relate to higher-order functional changes, as
well as help translate findings from animal models to the human realm. Finally, it may
introduce new measures for use as outcomes in clinical trials, a topic of intense debate
at the previous Workshop.

On the other hand, the psychology and neuroscience of great thinking is virtually non-
existent. Better insights come from literature and biography. Relativity is said to have
occurred to Einstein as part of a process of mental simulation and imagination over a
number of neighborhood walks. Proust began his *Remembrances* when the perfume of a
madeleine over a lazy morning coffee triggered a wave of obsessional work.

Ordinary mortals also need time to think. If you're co-ordinating 16 projects,
supervising 8 graduate students and drafting 6 papers simultaneously, it's unlikely you
have much of it. Our Workshop under the warm Provencal sun is therefore designed to
provide one of the necessary conditions for great thinking – unfettered time.

Whilst great thoughts can of course come anytime anyplace, the Workshop produced
some contenders. Independently, two groups in Perth and Barcelona presented
tantalizing new evidence about gene-environment interaction related to APOE4, otherwise considered a dementia and cardiovascular risk factor.

David Bartres-Faz showed that TMS applied to the prefrontal cortex produces much more robust changes in functional connectivity in APOE4 positive elders than individuals without the E4 allele. Some of the resultant changes were an amplification of those associated with compensatory brain activity in successful ageing, and so may be of longitudinal benefit.

Ralph Martins presented how a general negative correlation between lifestyle-related physical activity and amyloid brain burden is much stronger in APOE4 carriers, suggesting that exercise may be a particularly effective intervention in those at most risk for incident dementia.

Suddenly APOE4 is seen in a new light. Perhaps not only is it a risk factor but also an opportunity to intervene more successfully. Perhaps we need to combine TMS and exercise as part of a prevention trial in only APOE4 positive elders?

Yet perhaps the neatest example of original thinking was provided by Anders Fjell who approached old data with fresh eyes. He took advantage of the ADNI dataset that is freely available to all researchers (http://adni.loni.ucla.edu). This wonderful resource means anyone with a decent computer can carry out a significant study on one of the world's largest brain MRI datasets at virtually no cost. Actually, not anyone. You need to know what to do with the data, and moreover, have an original idea.

Anders' concern was the thorny issue of the distinction between so-called normal ageing and Alzheimer's disease (AD). He had the clever idea to group older individuals into those with no evidence of long term cognitive decline or presence of AD biomarkers, typically understood to preclude the possibility of subclinical AD, versus various other phenotypes. He showed that longitudinal atrophy of the entorhinal cortex, one of the areas first targeted by AD, was detectable even in these 'super stable' cognitively-intact elders. Accelerated atrophy in those brain area often linked with AD does not uniquely signify neurodegenerative illness, but instead may also comprise 'normal aging' – whatever that is.

To be clear there is nothing wrong with big data-hungry science. Indeed, there has in recent years been an astonishing growth in generation of truly enormous datasets. Yet one forceful message from the Workshop was that along with truckloads of data we need a commensurate clearing out of mental space. Great thoughts are ultimately the best type of data reduction procedure.
References


