

Reproducibility in Computationally-Intensive Behavioral Research

Rick O. Gilmore
8/22/2017

Preliminaries



Eunice Kennedy Shriver
National Institute of
Child Health and
Human Development



**Alfred P. Sloan
FOUNDATION**

Overview

- The reproducibility "crisis"
- The "crisis"" in psychology
- "Big data" behavioral science is computationally intensive
- Let's not waste a "good" crisis

The reproducibility "crisis"

Is there a reproducibility crisis?

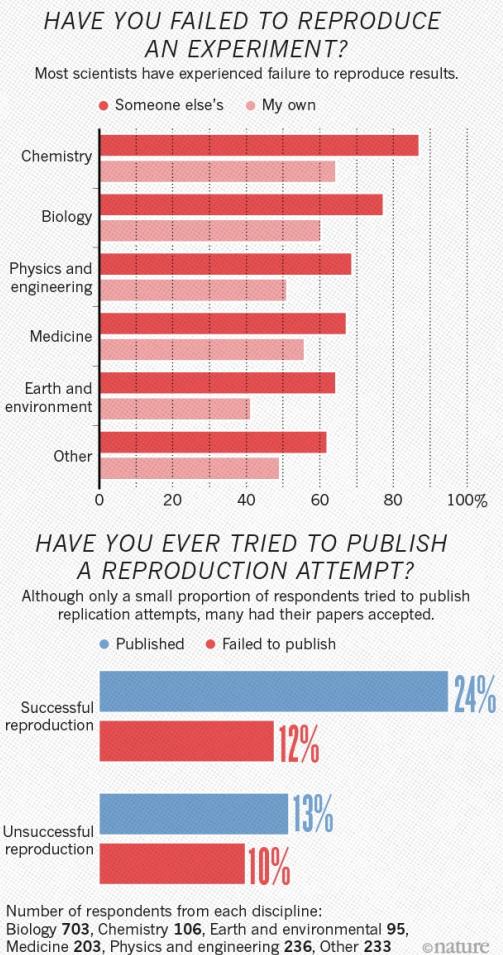
- Yes, a significant crisis
- Yes, a slight crisis
- No crisis
- Don't know

IS THERE A REPRODUCIBILITY CRISIS?



©nature

([Baker, 2016](#))



(Baker, 2016)

What does "reproducibility" mean?

Methods reproducibility

- Enough details about materials & methods recorded (& reported)
- Same results with same materials & methods

(Goodman et al., 2016)

Results reproducibility

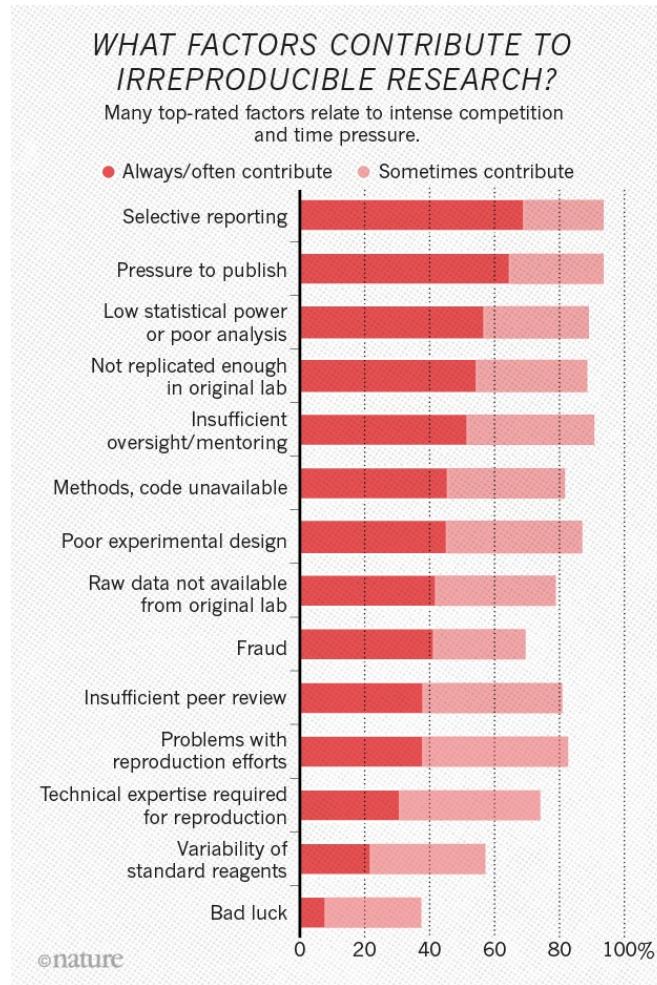
- Same results from independent study

([Goodman et al., 2016](#))

Inferential reproducibility

- Same inferences from one or more studies or reanalyses

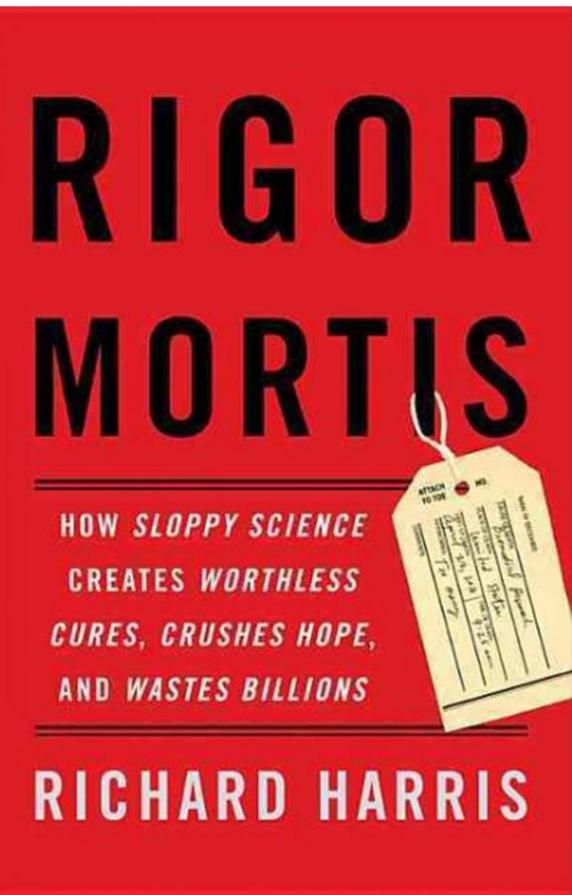
([Goodman et al., 2016](#))



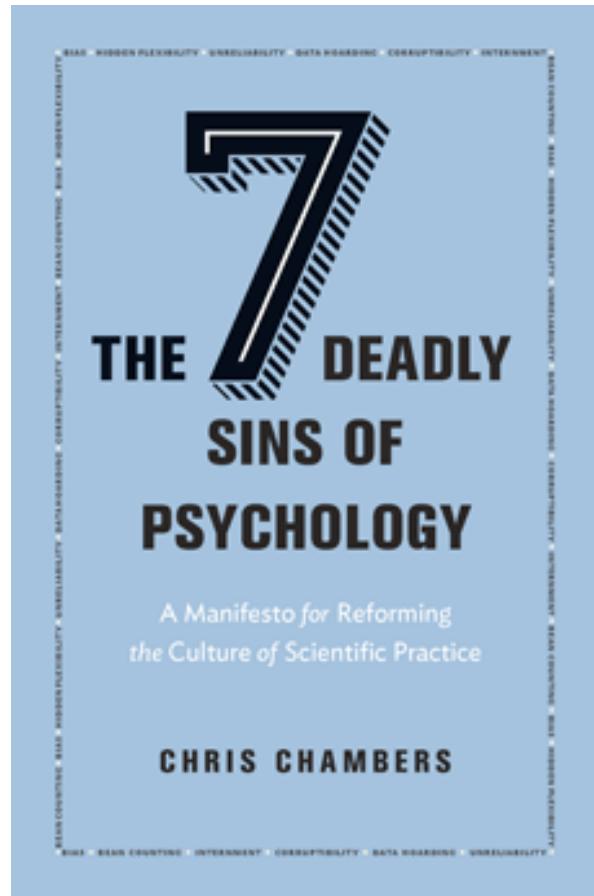
(Baker, 2016)

Reproducibility crisis

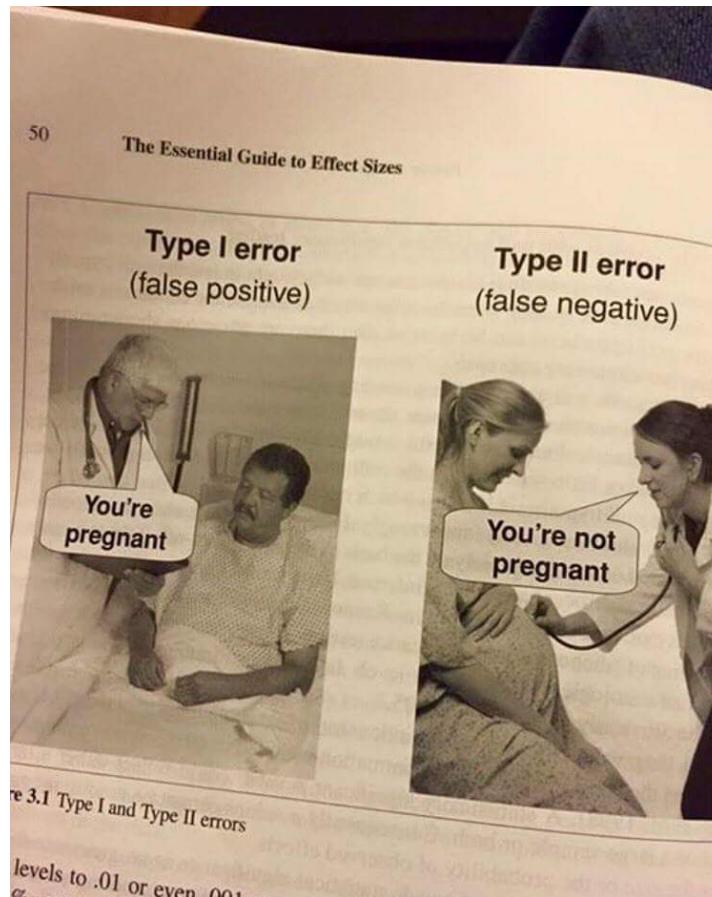
- Not just psychology
- "Hard" sciences, too
- Data collection to statistical analysis to reporting to publishing



The crisis in psychology

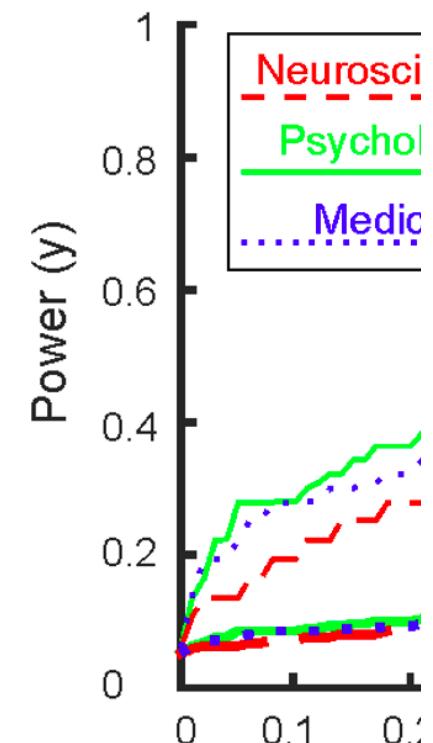
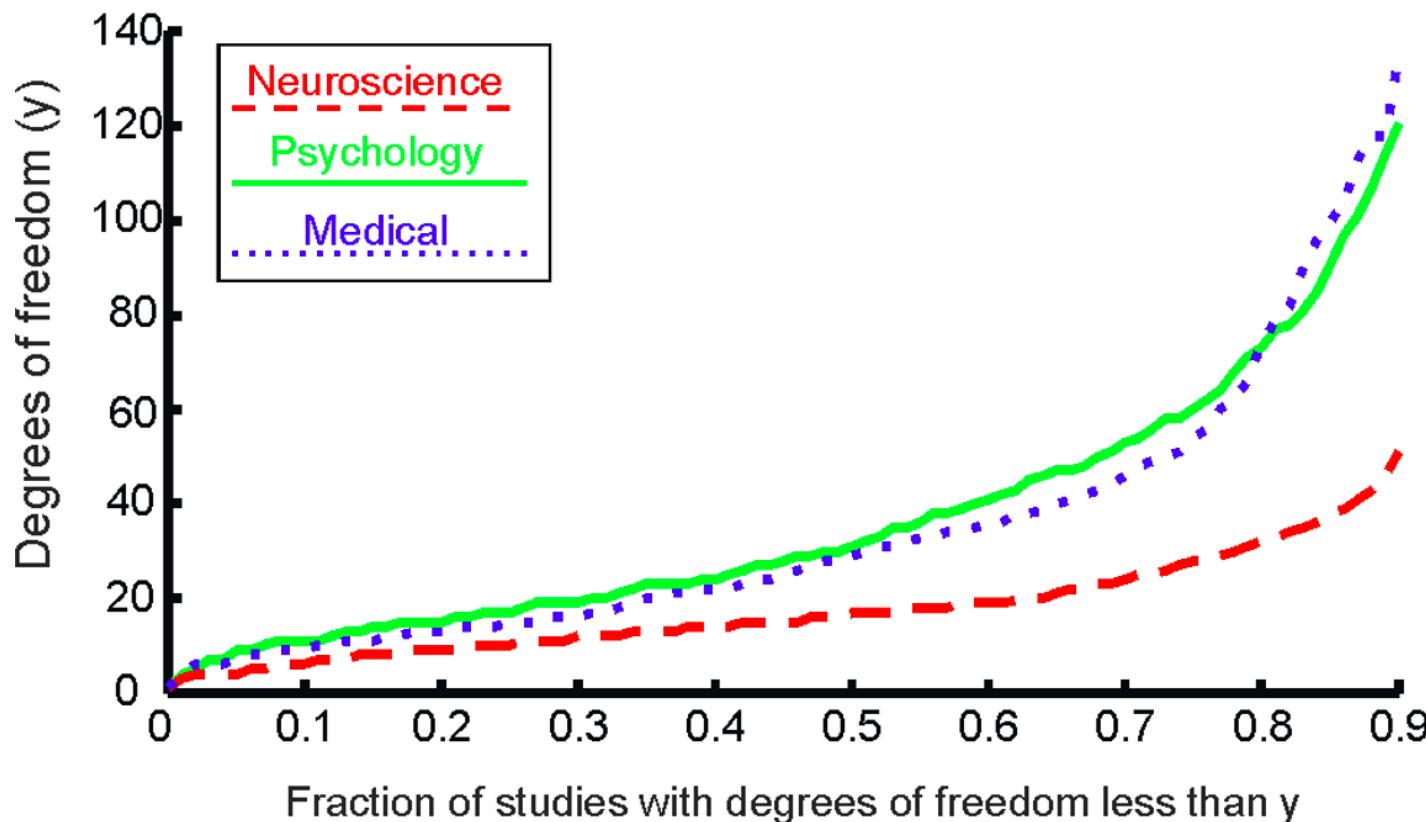


The sin of unreliability



Studies are underpowered

A. Degrees of freedom in subfields

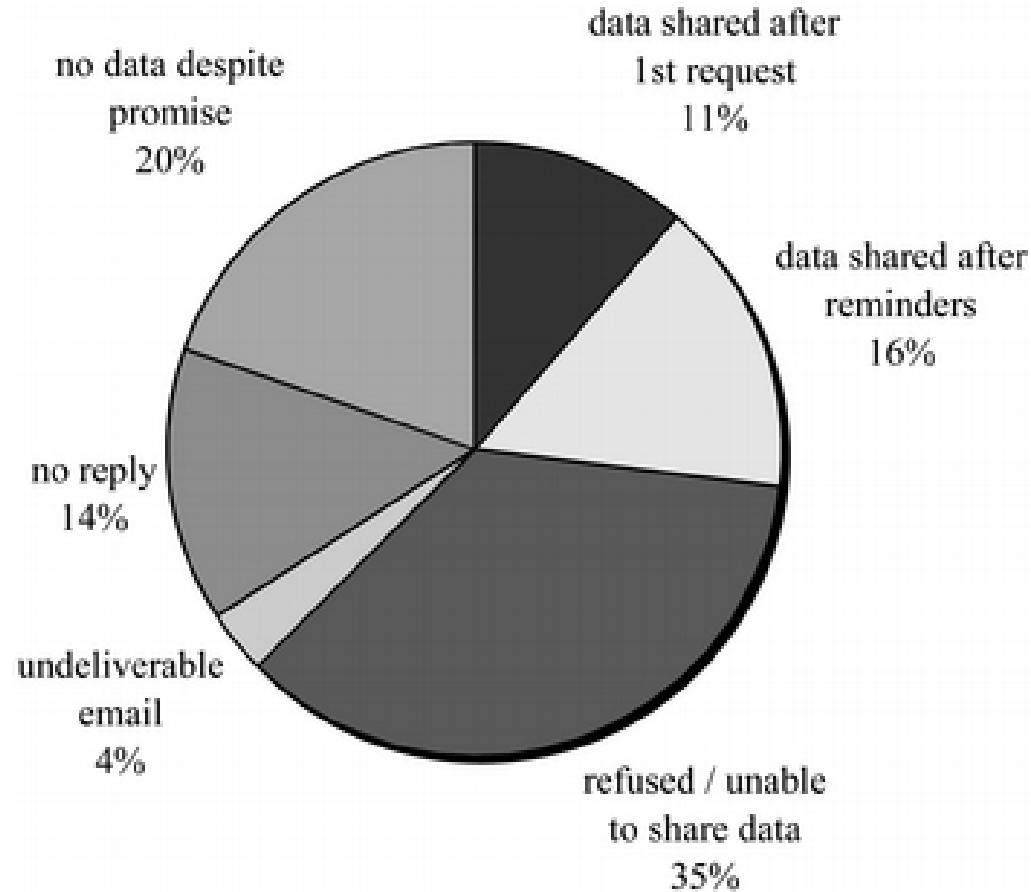


(Szucs & Ioannides, 2017)

"Assuming a realistic range of prior probabilities for null hypotheses, false report probability is likely to exceed 50% for the whole literature."

(Szucs & Ioannides, 2017)

The sin of hoarding...



(Wicherts et al., 2006)

The sin of corruptibility...

REPORT

When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour¹, Donald P. Green²

[+ See all authors and affiliations](#)

Science 12 Dec 2014;
Vol. 346, Issue 6215, pp. 1366-1369
DOI: 10.1126/science.1256151

[Article](#)

[Figures & Data](#)

[Info & Metrics](#)

[eLetters](#)

 [PDF](#)

You are currently viewing the abstract.

[View Full Text](#)



This article has been retracted. Please see:
[Is retracted by - June 05, 2015](#)

[\(LaCour & Green, 2014\)](#)

Flawed science: The fraudulent research practices of social psychologist Diederik Stapel

Levelt Committee

Noort Committee

Drenth Committee

Science

AAAS

Home

News

Journals

Topics

Careers

Latest News

ScienceInsider

ScienceShots

Sifter

From the Magazine

About News

Quizzes

SHARE



7

Harvard Psychology Researcher Committed Fraud, U.S. Investigation Concludes



By Siri Carpenter | Sep. 6, 2012, 4:57 PM

The sin of bias...

Bem, D.J. (2011). Experimental evidence for anomalous retroactive influences on cognition and affect. , (3), 407-425.

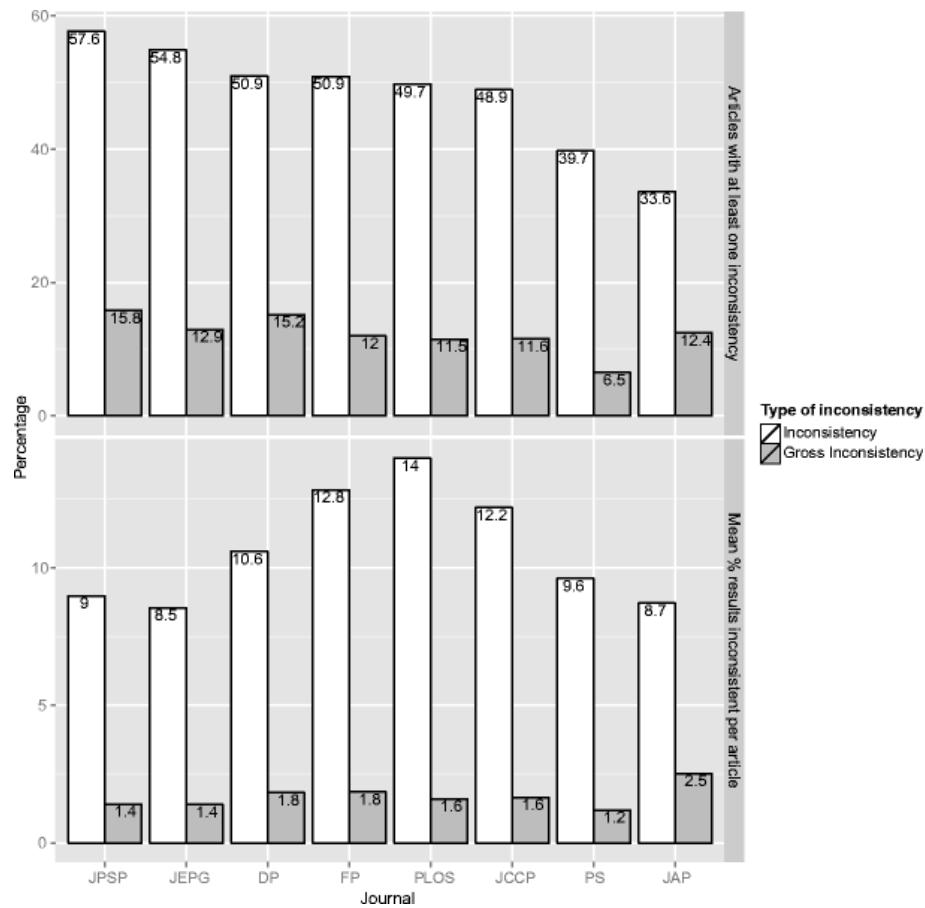
"This article reports 9 experiments, involving more than 1,000 participants, that test for retroactive influence by "time-reversing" well-established psychological effects so that the individual's responses are obtained before the putatively causal stimulus events occur."



"We argue that in order to convince a skeptical audience of a controversial claim, one needs to conduct strictly confirmatory studies and analyze the results with statistical tests that are conservative rather than liberal. We conclude that Bem's values do not indicate evidence in favor of precognition; instead, they indicate that experimental psychologists need to change the way they conduct their experiments and analyze their data."

([Wagenmakers et al., 2011](#))

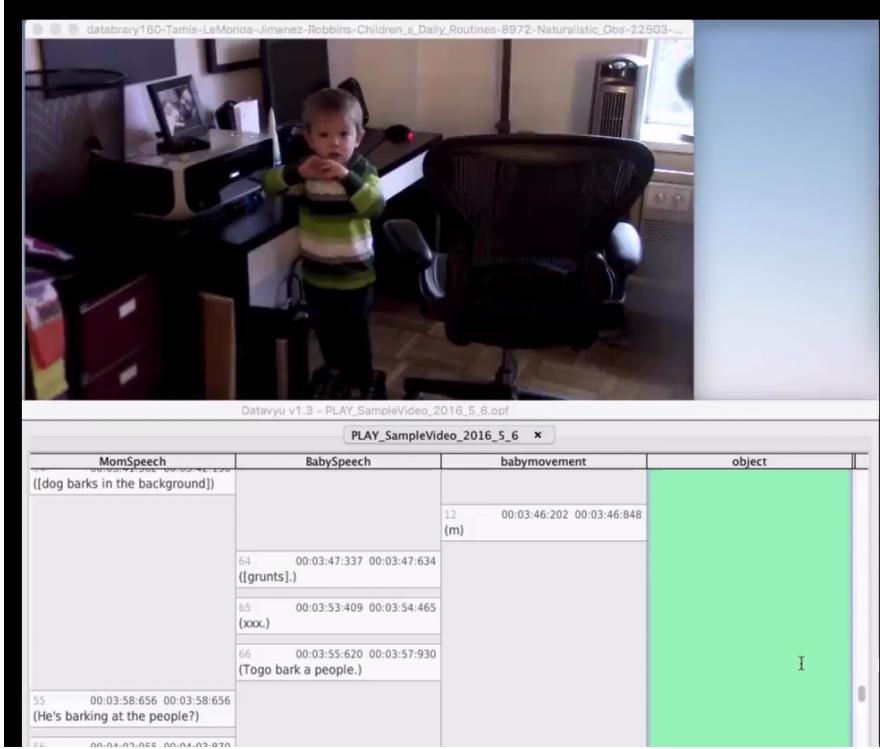
The sin of hurrying...



(Nuijten et al., 2015)

In our defense...

Behavior multidimensional

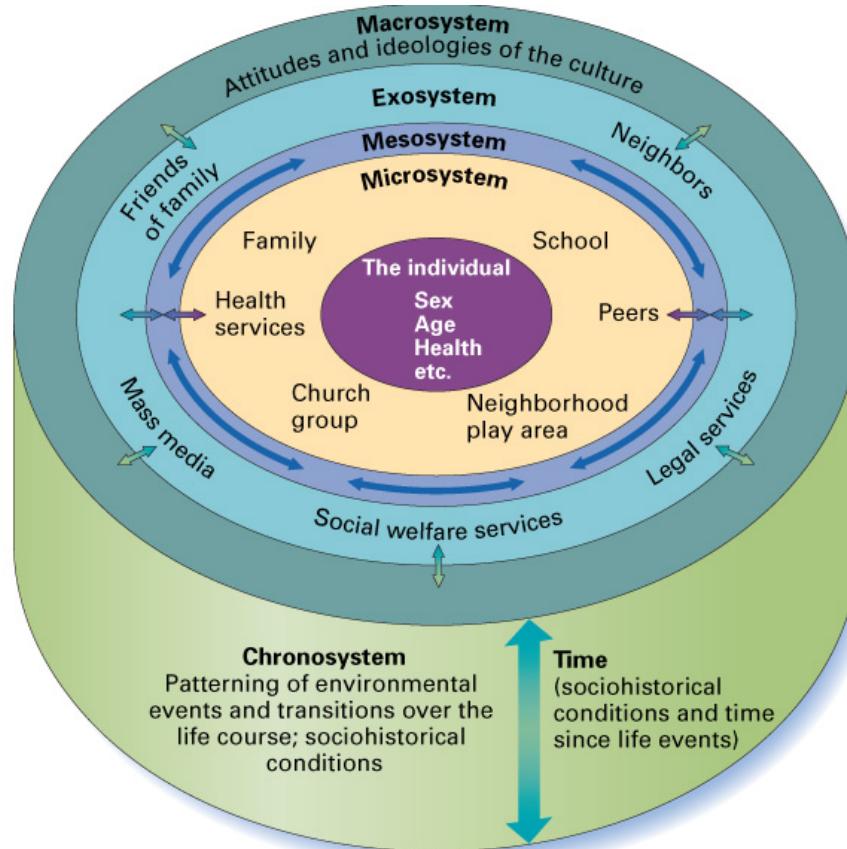


The image shows a screenshot of the Datavyu v1.3 software interface. At the top, there is a video frame showing a young child standing in a room with a computer monitor and other furniture. Below the video frame is a timeline table titled "PLAY_SampleVideo_2016_5_6". The table has four columns: "MomSpeech", "BabySpeech", "babymovement", and "object". The "MomSpeech" column contains several rows of text, some of which are highlighted in green. The "BabySpeech" column contains a single row of text. The "babymovement" column contains a single row of text. The "object" column contains a large green rectangular area. The timeline table also includes a header row with column names and a footer row with numerical values.

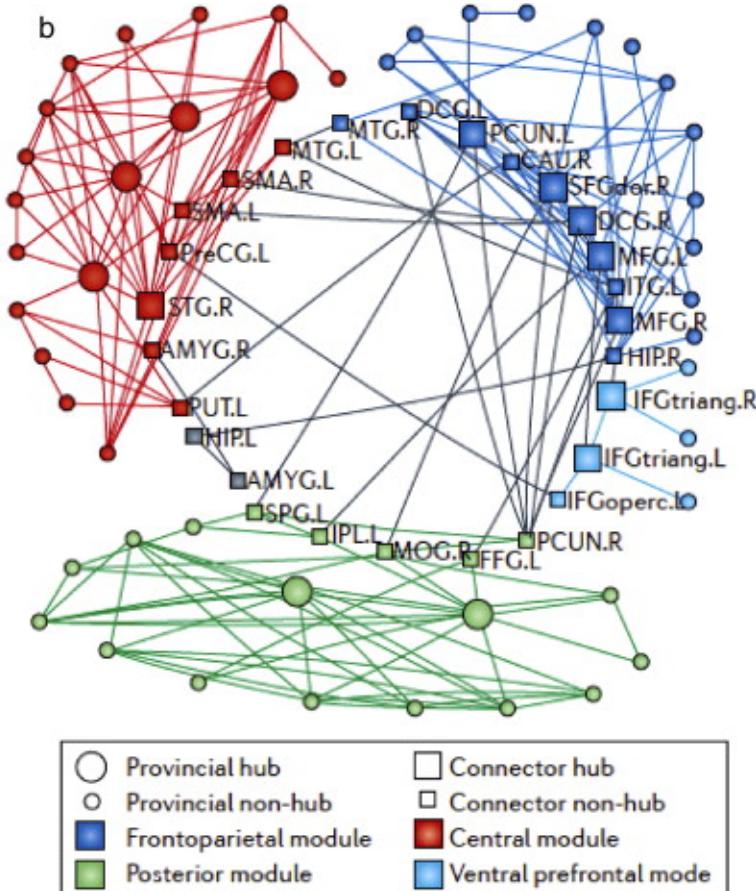
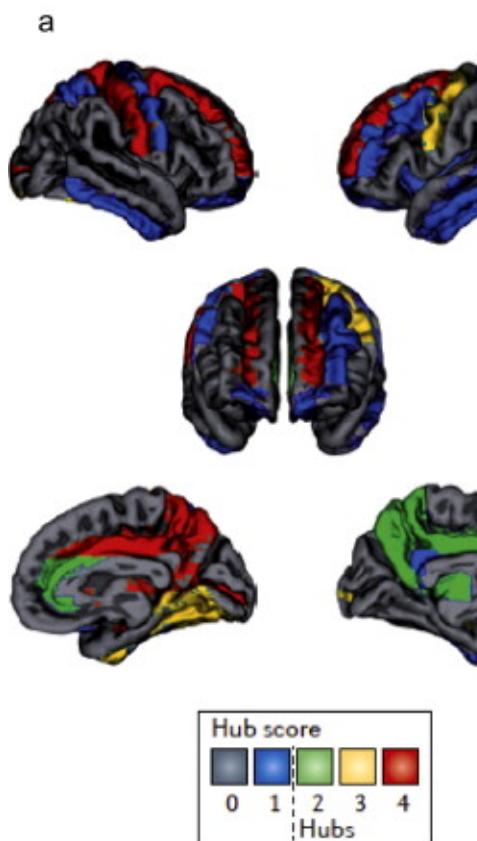
MomSpeech	BabySpeech	babymovement	object
(dog barks in the background)			
		12 00:03:46:202 00:03:46:848 (m)	
	64 00:03:47:337 00:03:47:634 ([grunts.])		
	65 00:03:53:409 00:03:54:465 (xxx.)		
	66 00:03:55:620 00:03:57:930 (Togo bark a people.)		I
55 00:03:58:656 00:03:58:656 (He's barking at the people?)			

[\(Adolph et al., 2016\)](#)

Embedded in networks

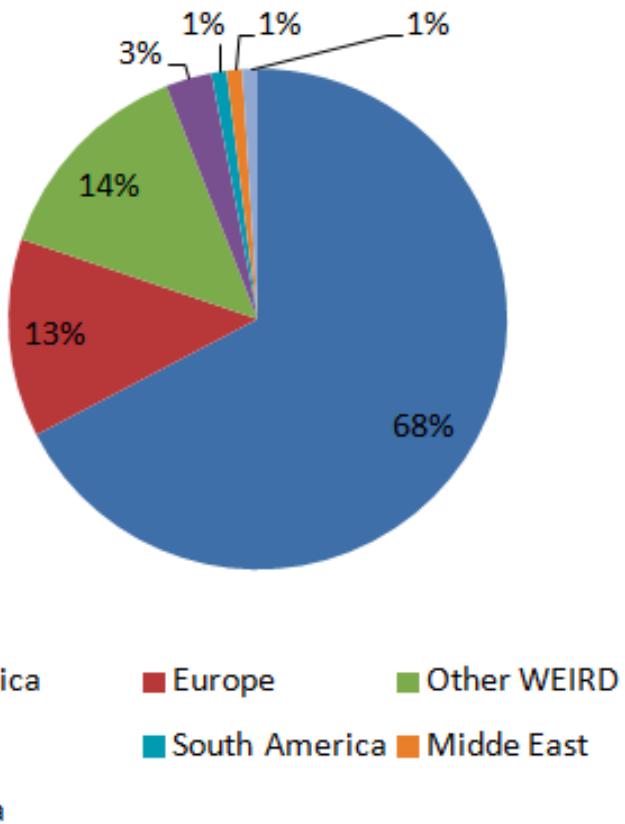


Copyright ©The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



Humans are diverse

- But much (lab-based) data collected are from Western, Educated Industrialized, Rich, Democratic (WEIRD) populations [Henrich et al., 2010](#)

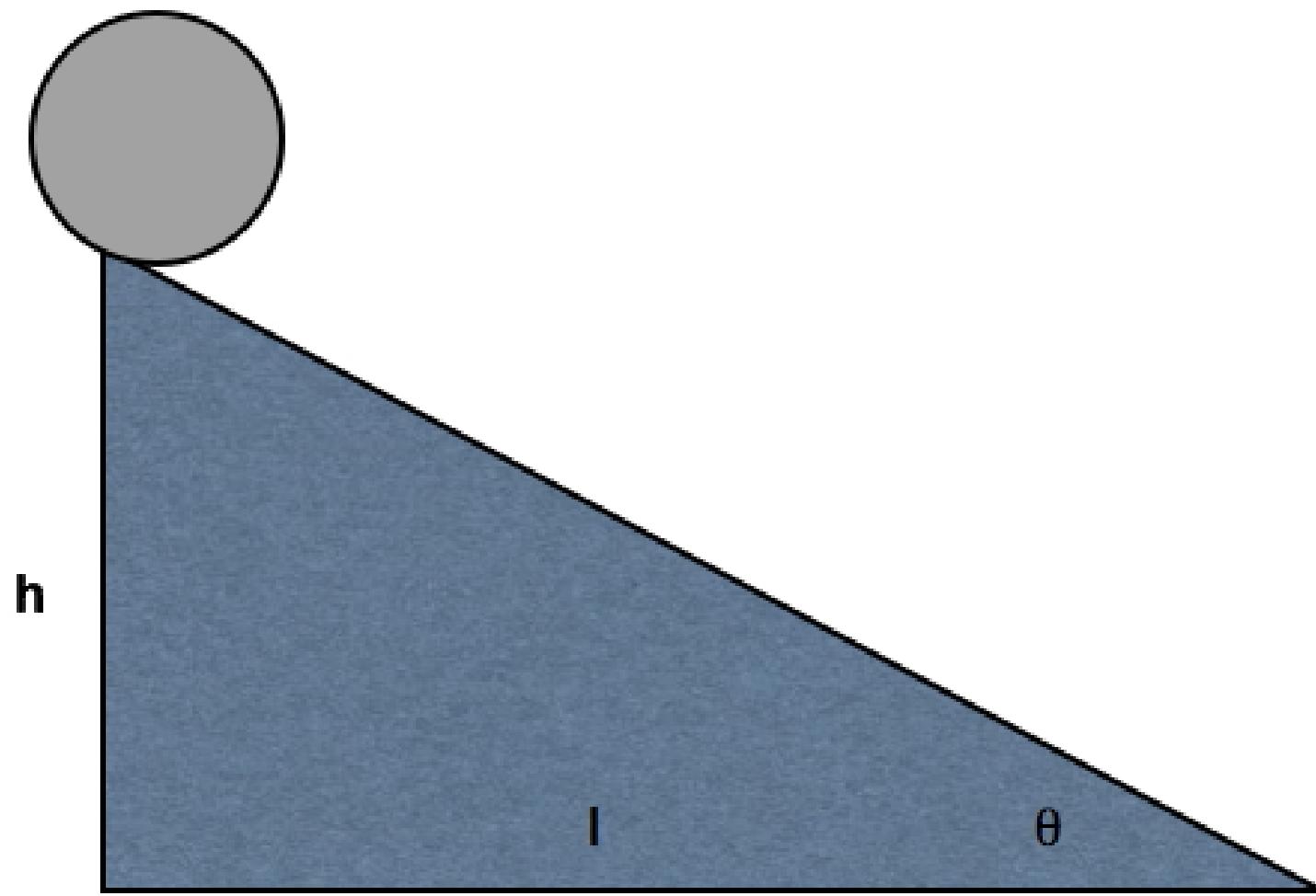


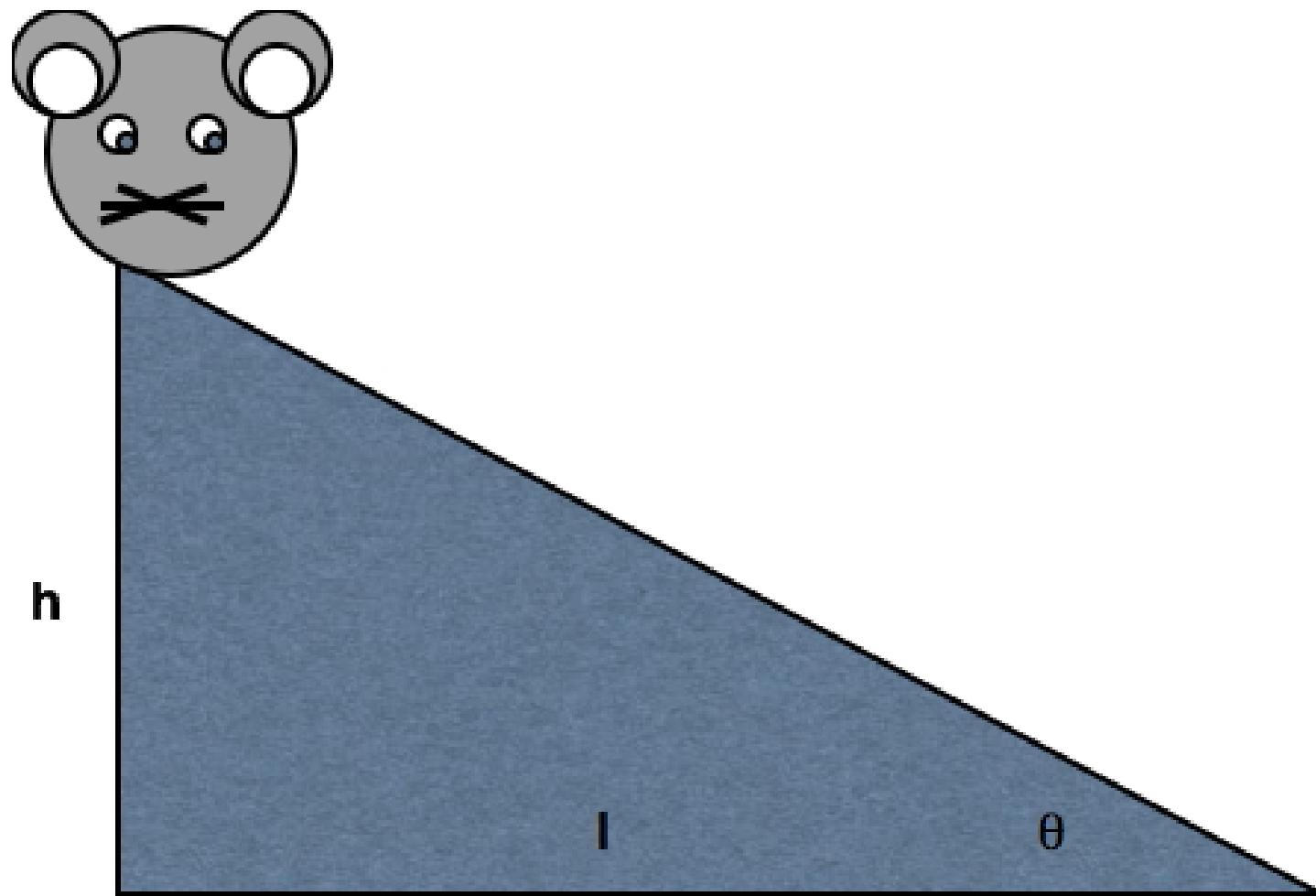
<http://www.evoanth.net/2015/01/06/evolutionary-psychology-has-problems-and-it-isnt-getting-better/>

Data sensitive, hard(er) to share

- Protect participant's identities
- Protect from harm/embarrassment
- Anonymize (effective?) or get permission

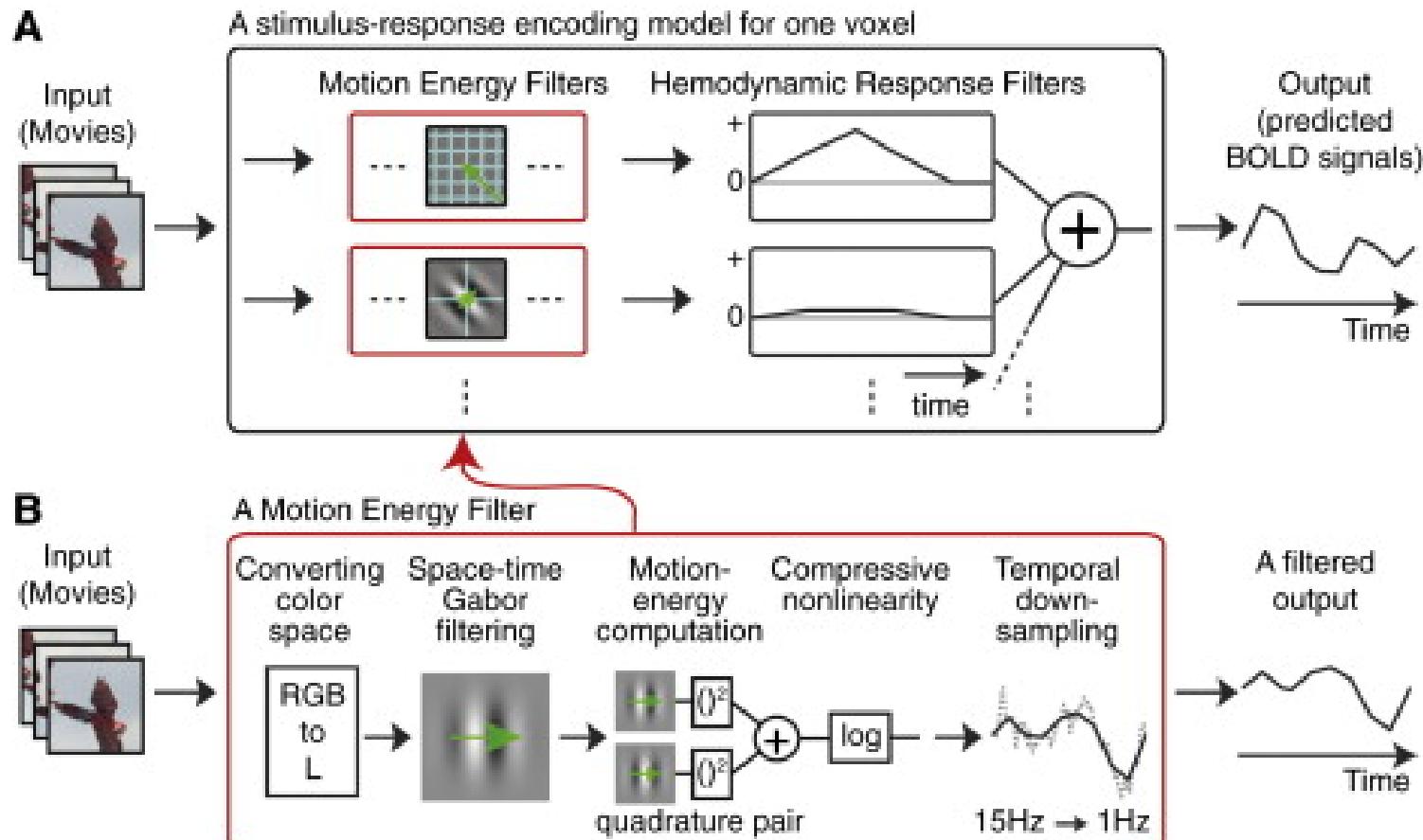
Psychology is harder than physics



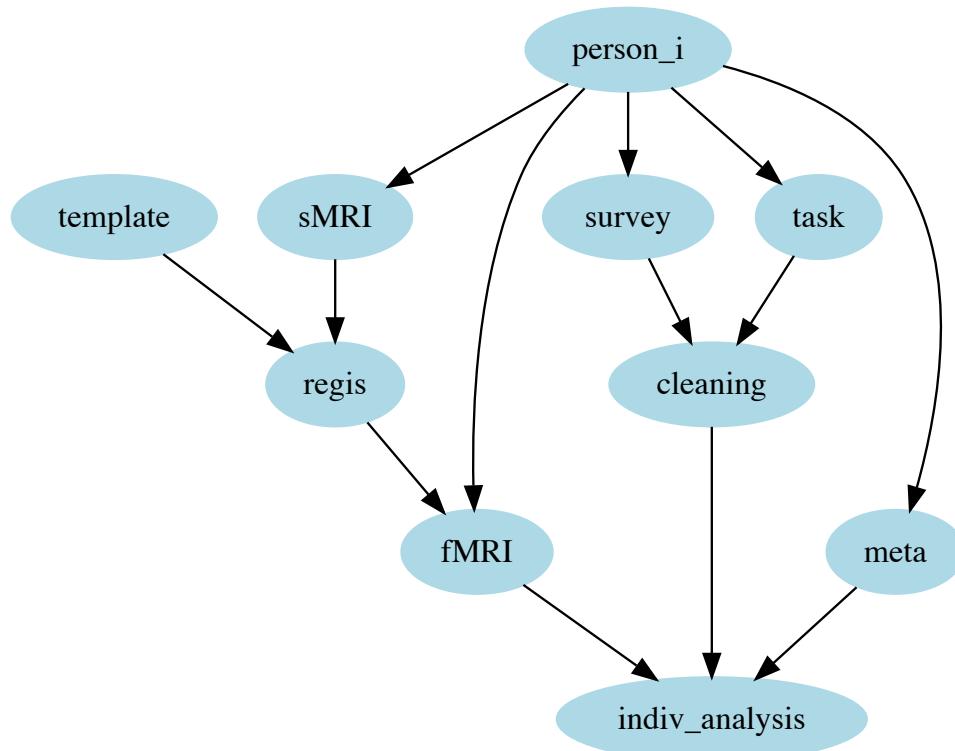


Big data computation in psychological science

"Mind-reading" in fMRI



(Nishimoto et al., 2011)

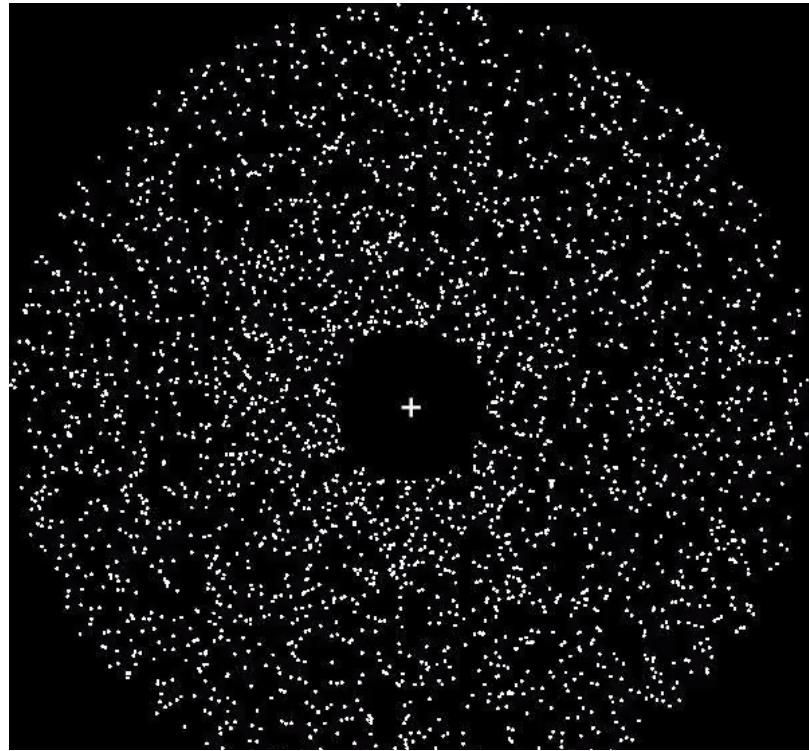


A personal example

- How does vision develop?
- - Input +
 - Visually-guided action
- - (eye/brain/body) development

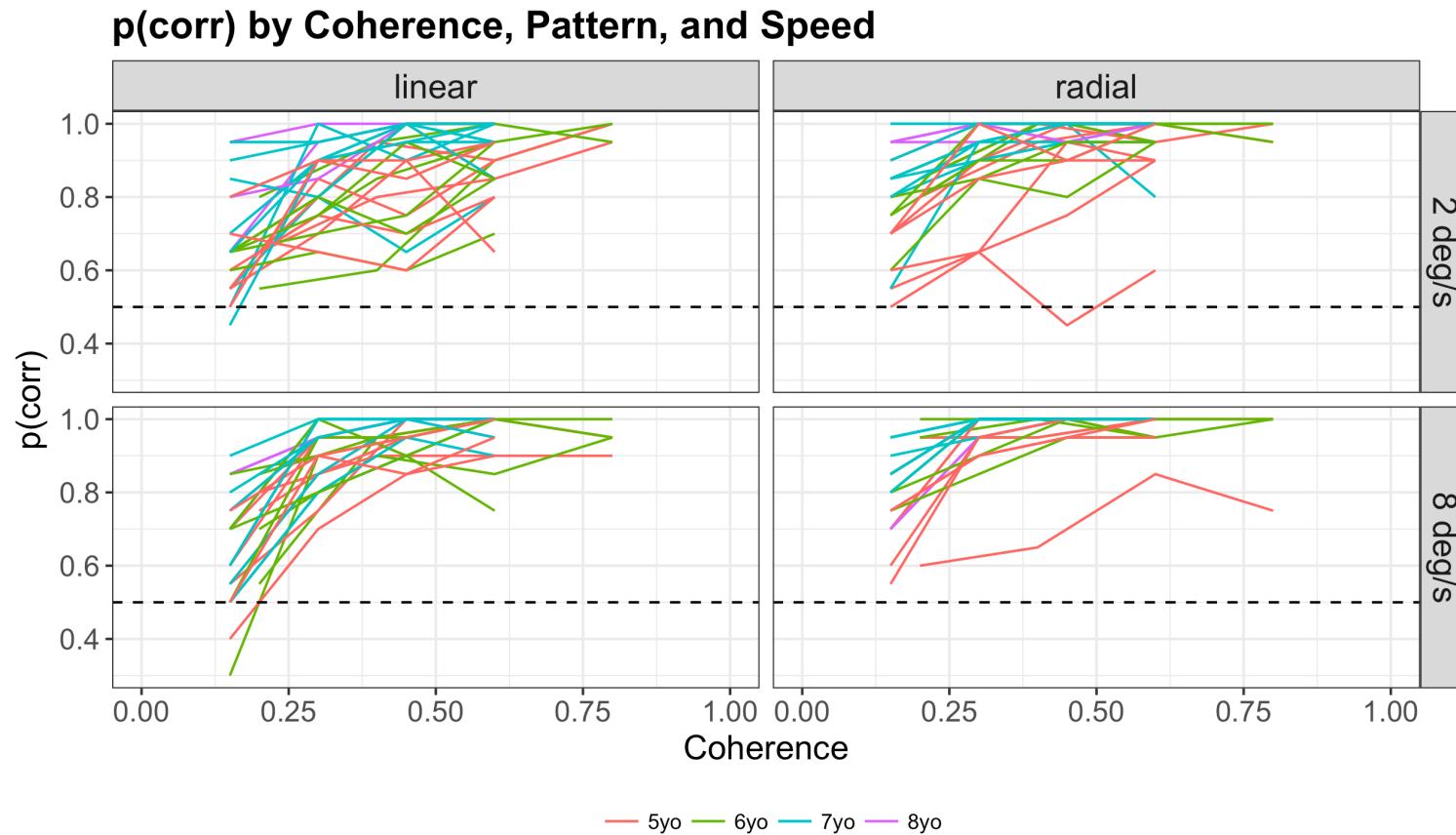
Measure (in the lab)

- Behavioral sensitivity
- Brain responses
- At different ages



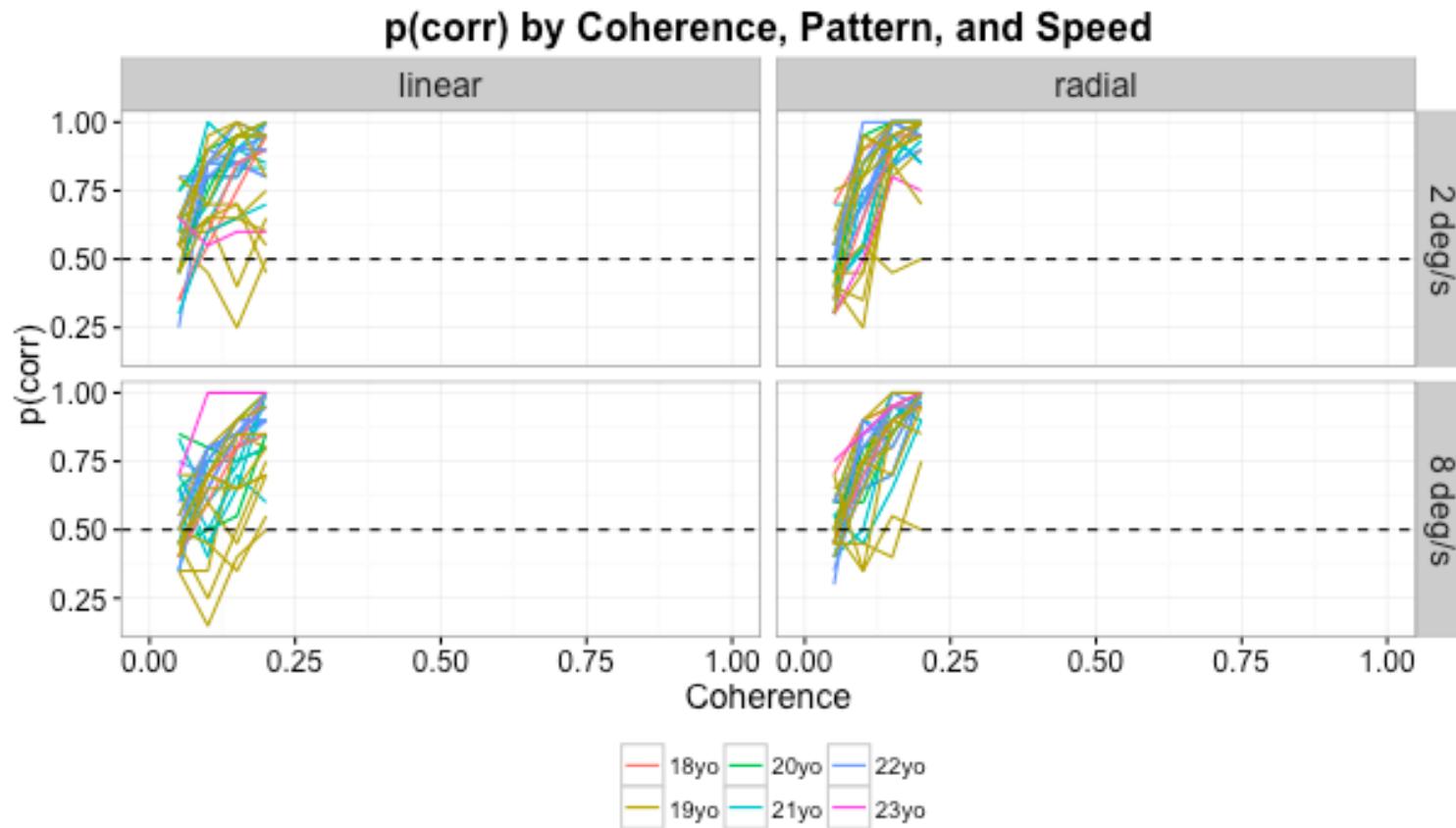
(Gilmore, 2014)

Children's behavior



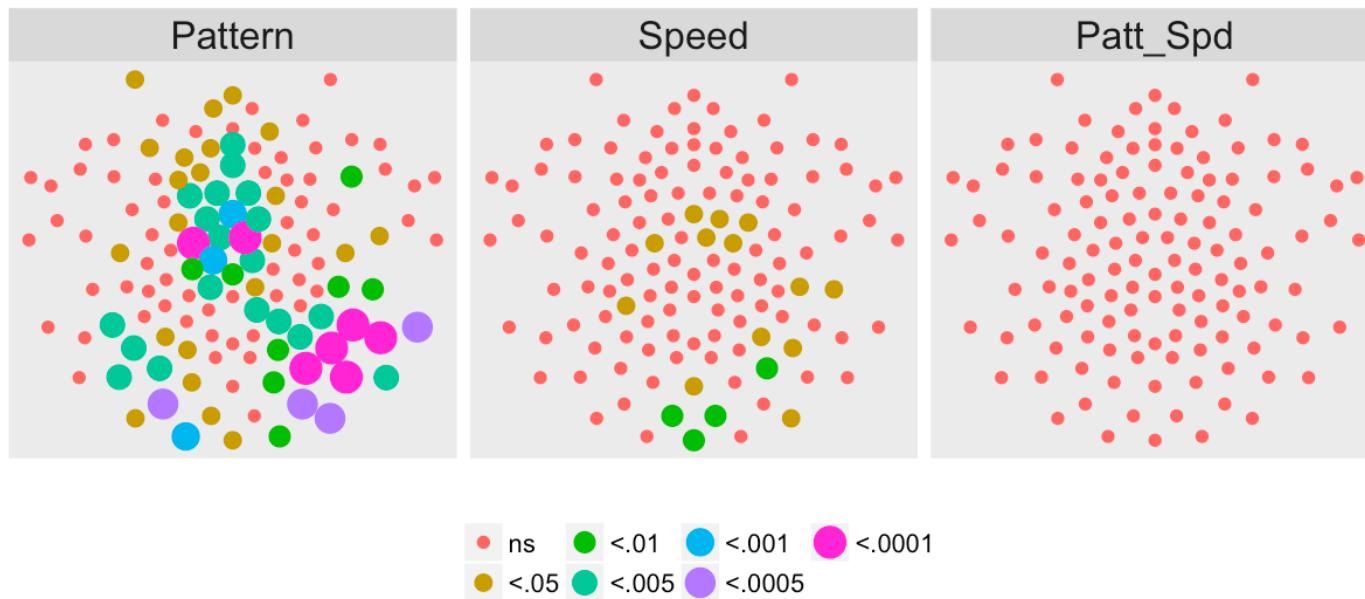
([Adamiak et al., 2015](#))

Adults' behavior



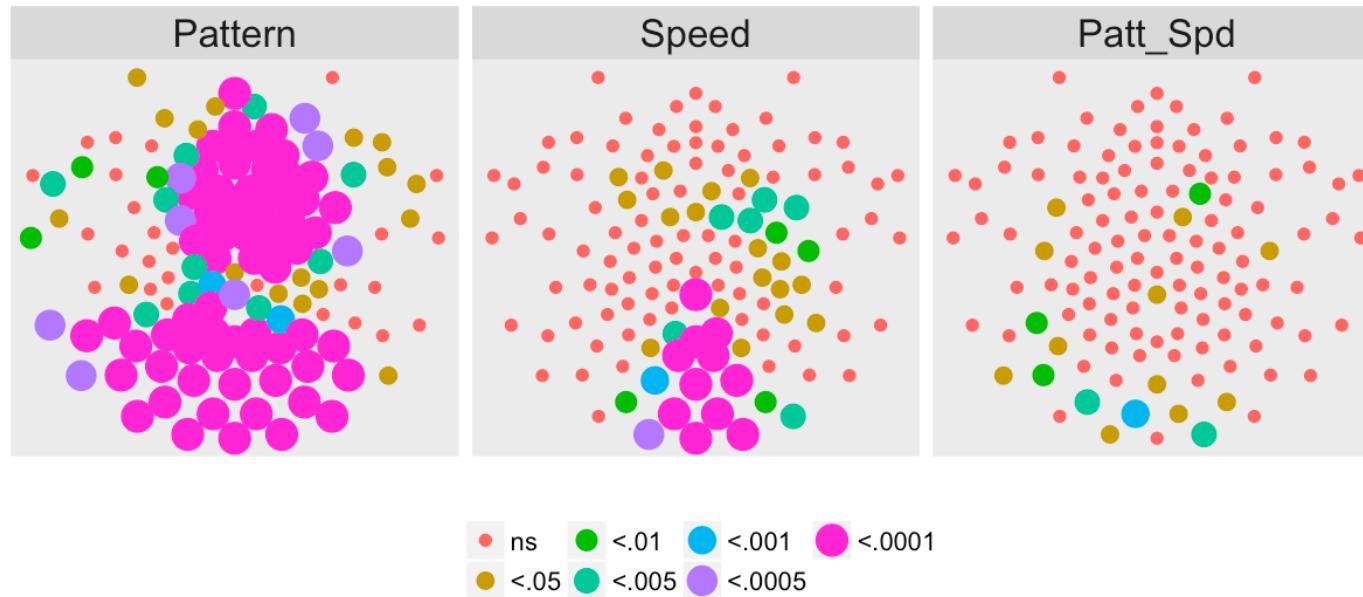
([Adamiak et al., 2015](#))

Children's brain responses



([Gilmore et al., 2015](#))

Adults' brain responses



([Gilmore et al., 2015](#))

But, what's the input? The *real* input?



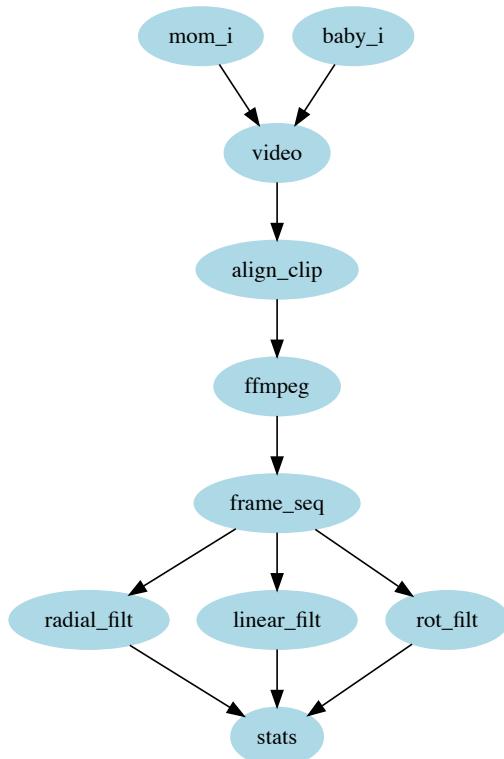
([Gilmore et al., 2015](#))



([Gilmore et al., 2015](#))



(Adolph, 2015)



Frame-by-frame video analysis



(Jayaraman et al., 2015)

Flow 00001->00002



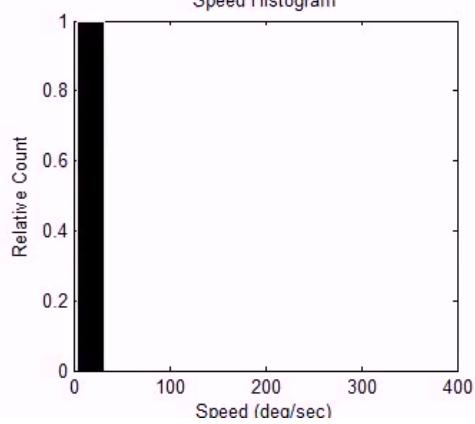
Flow Pattern Match

-1- -1- 2 15
-- -- !! !!

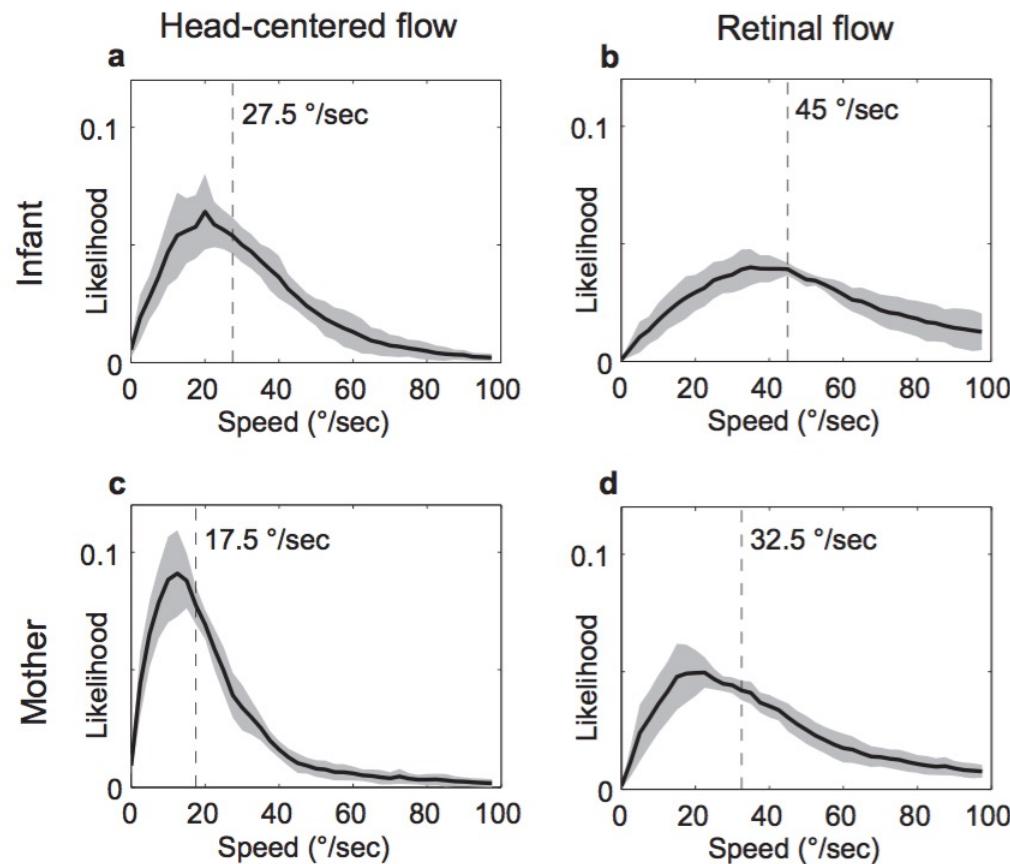
Flow 00001->00002



Speed Histogram



Findings



(Raudies & Gilmore, 2014)

Findings

- Infant (passengers) experience faster visual speeds than mother
- Controlling for speed of locomotion, environment
- Motion "priors" for infants ≠ mothers

Are "fast" flow speeds common?



(Jayaraman et al., 2015)

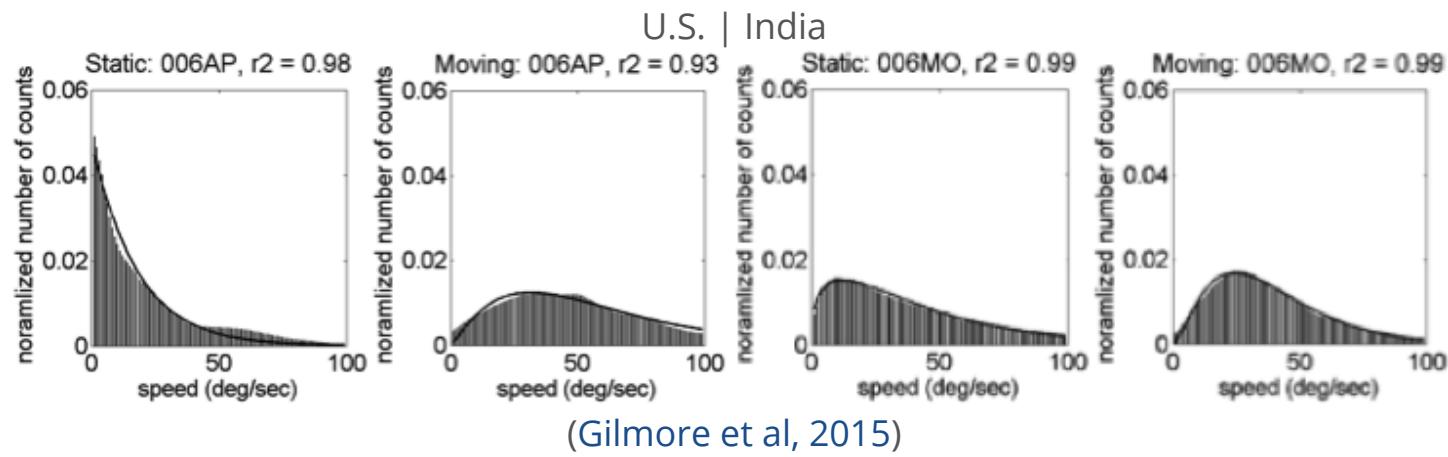


([Gilmore et al, 2015](#))

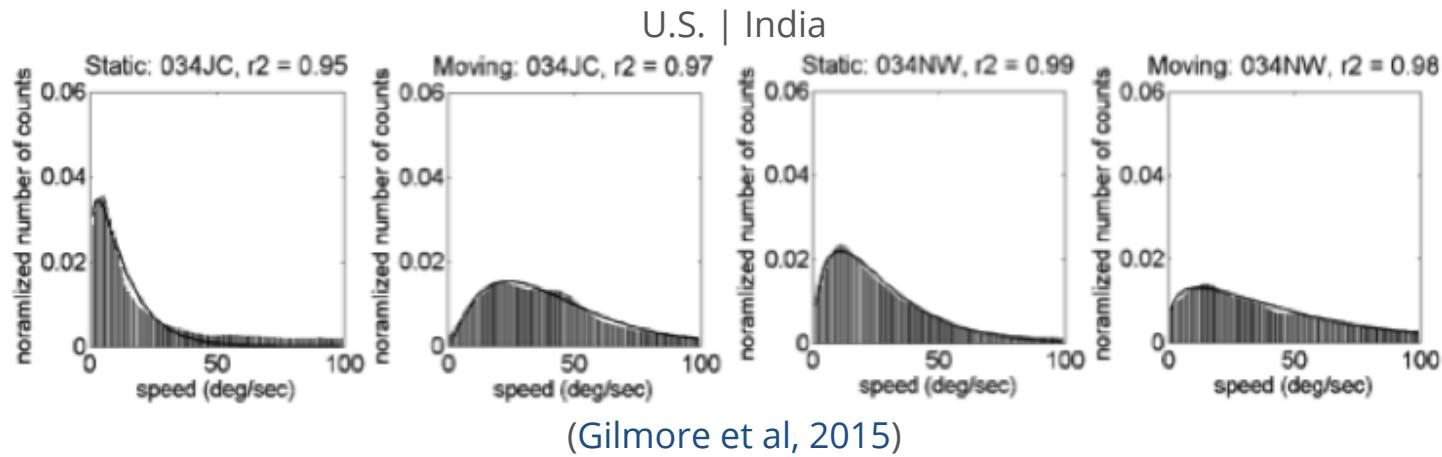
Country	Females	Males	Age (wks)	Coded video Hrs
India	17	13	3-63	3.1 (0.5-6.0)
U.S.	15	19	4-62	4.6 (0.2-7.6)

(Gilmore et al, 2015)

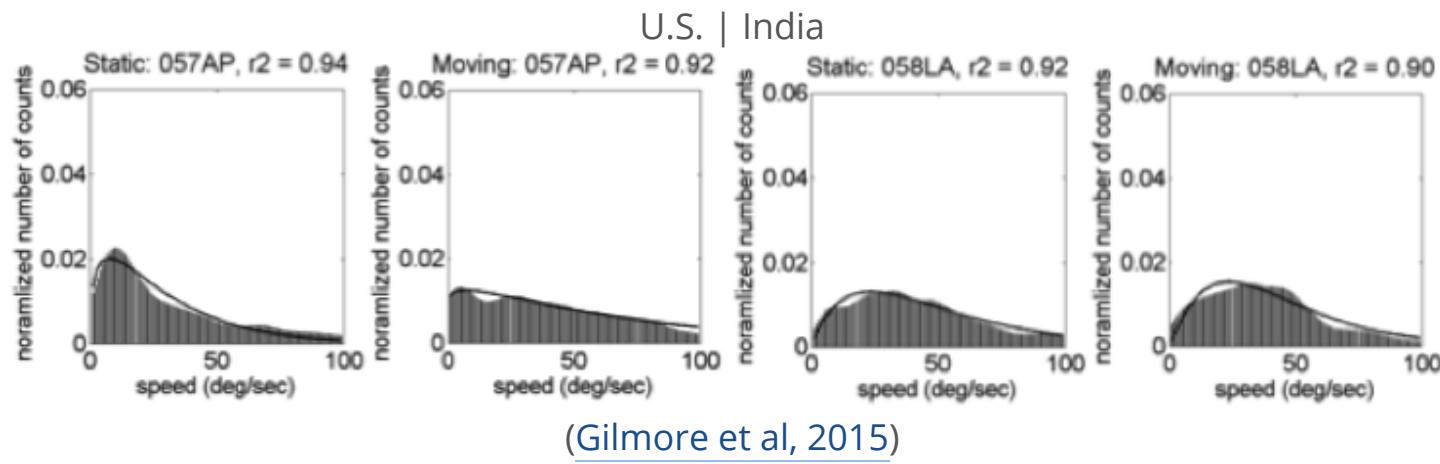
Motion speeds - 6 weeks



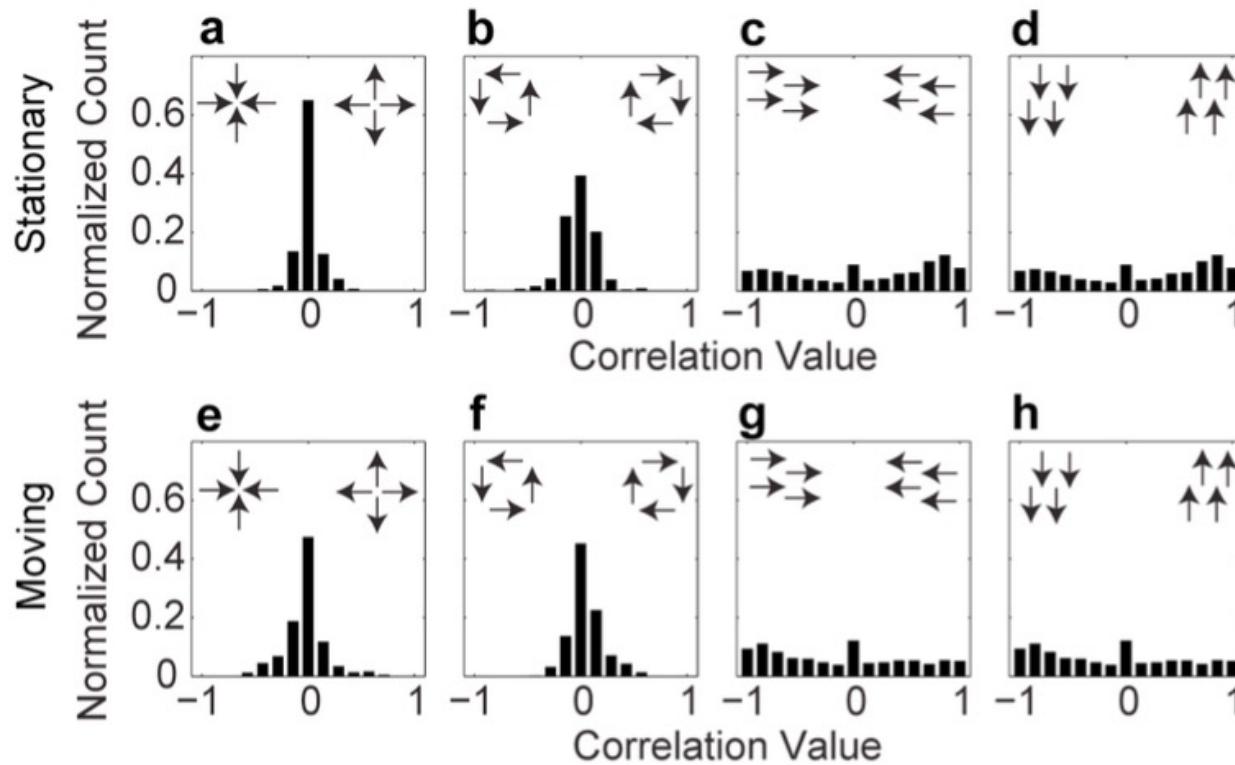
Motion speeds – 34 weeks



Motion speeds – 58 weeks



Linear > radial patterns



([Gilmore et al, 2015](#))

Simulating developmental change

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \frac{1}{z} \begin{pmatrix} -f & 0 & x \\ 0 & -f & y \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} + \frac{1}{f} \begin{pmatrix} xy & -(f^2 + x^2) & fy \\ f^2 + y^2 & -xy & -fy \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix}$$

Geometry of environment/observer: (x, y, z)

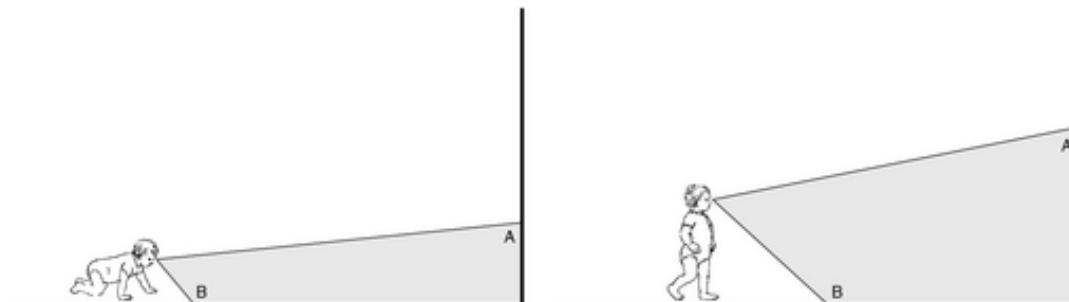
Translational speed: (v_x, v_y, v_z)

Rotational speed: $(\omega_x, \omega_y, \omega_z)$

Retinal flow: (\dot{x}, \dot{y})

Parameters For Simulation

Parameter	Crawling Infant	Walking Infant
Eye height	0.30 m	0.60 m
Locomotor speed	0.33 m/s	0.61 m/s
Head tilt	20 deg	9 deg

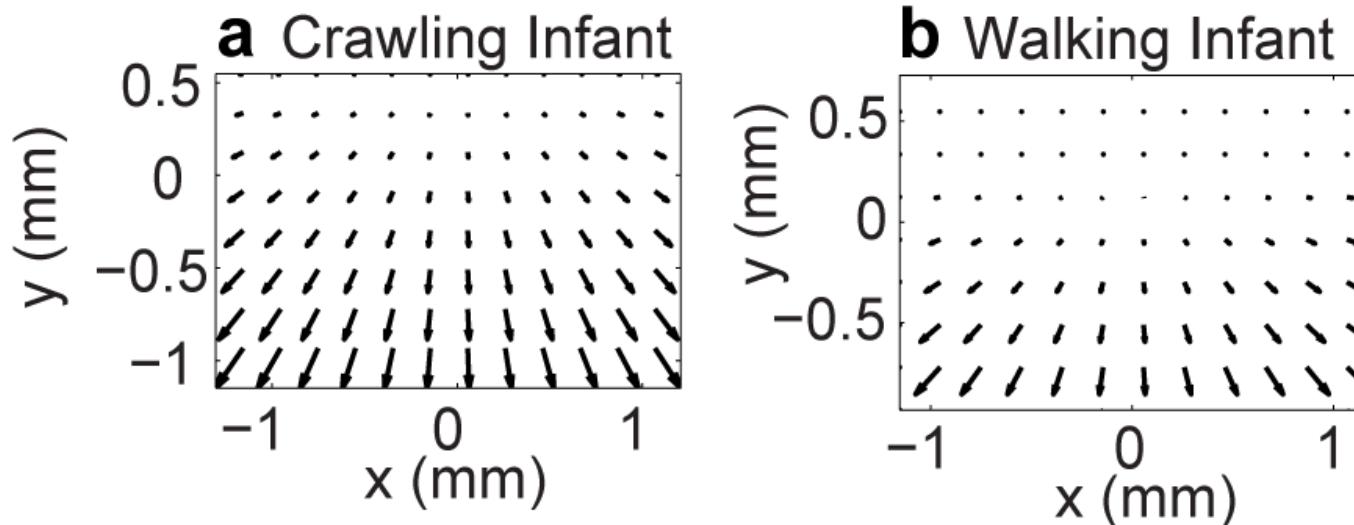


[\(Kretch et al., 2014\)](#)

Parameters for Simulation

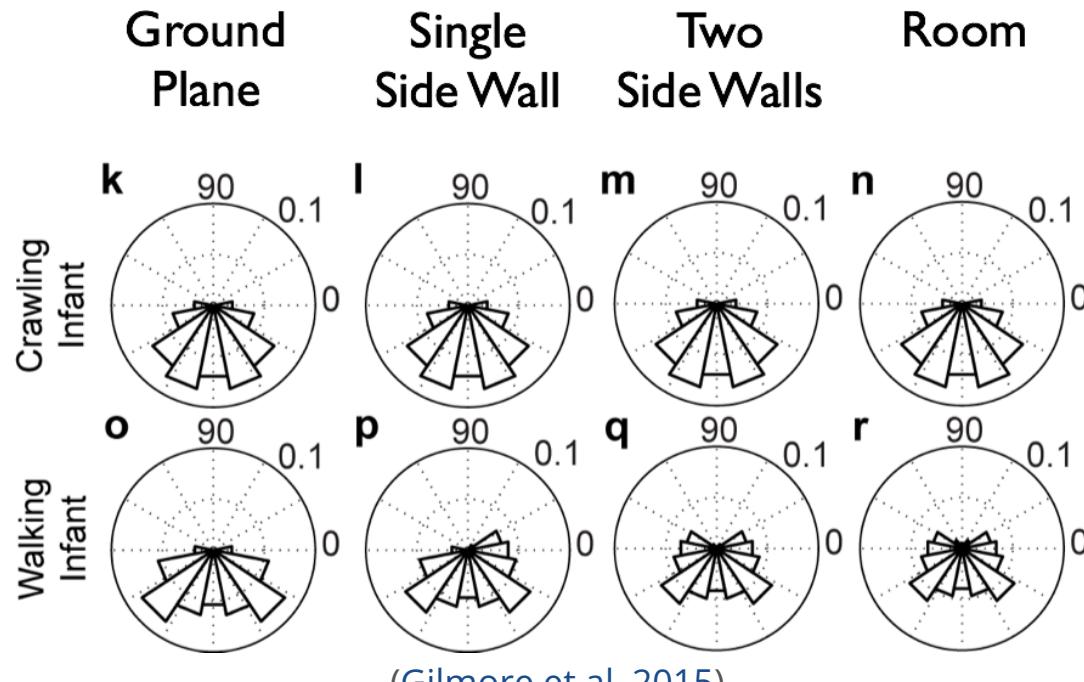
Geometric Feature	Distance
Side wall	+/- 2 m
Side wall height	2.5 m
Distance of ground plane	32 m
Field of view width	60 deg
Field of view height	45 deg

Simulating Flow Fields



([Gilmore et al, 2015](#))

Flow Direction Distributions by Geometry & Posture



Simulated Flow Speeds (m/s)

Type of Locomotion	Ground Plane	Room	Side Wall	Two Walls
Crawling	14.41	14.42	14.43	14.62
Walking	9.38	8.56	7.39	9.18

Essentials for computationally intensive psychological research

- Computational resources
- Technical expertise

Create reproducible workflows

Kitzes, J., Turek, D., & Deniz, F. (Eds.). (2018). The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences. Oakland, CA: University of California Press. [E-book.](#)

Share materials, code, raw data



Databrary is a video data library for developmental science.
Share videos, audio files, and related metadata. Discover more, faster.

593 investigators 248 affiliates 332 institutions

401 total volumes 75 shared volumes 43,193 files 8,151 hours of recordings 8,595 participants



Your Profile

Manage your data and authorizations.



Browse Data



Browse People

databrary.org

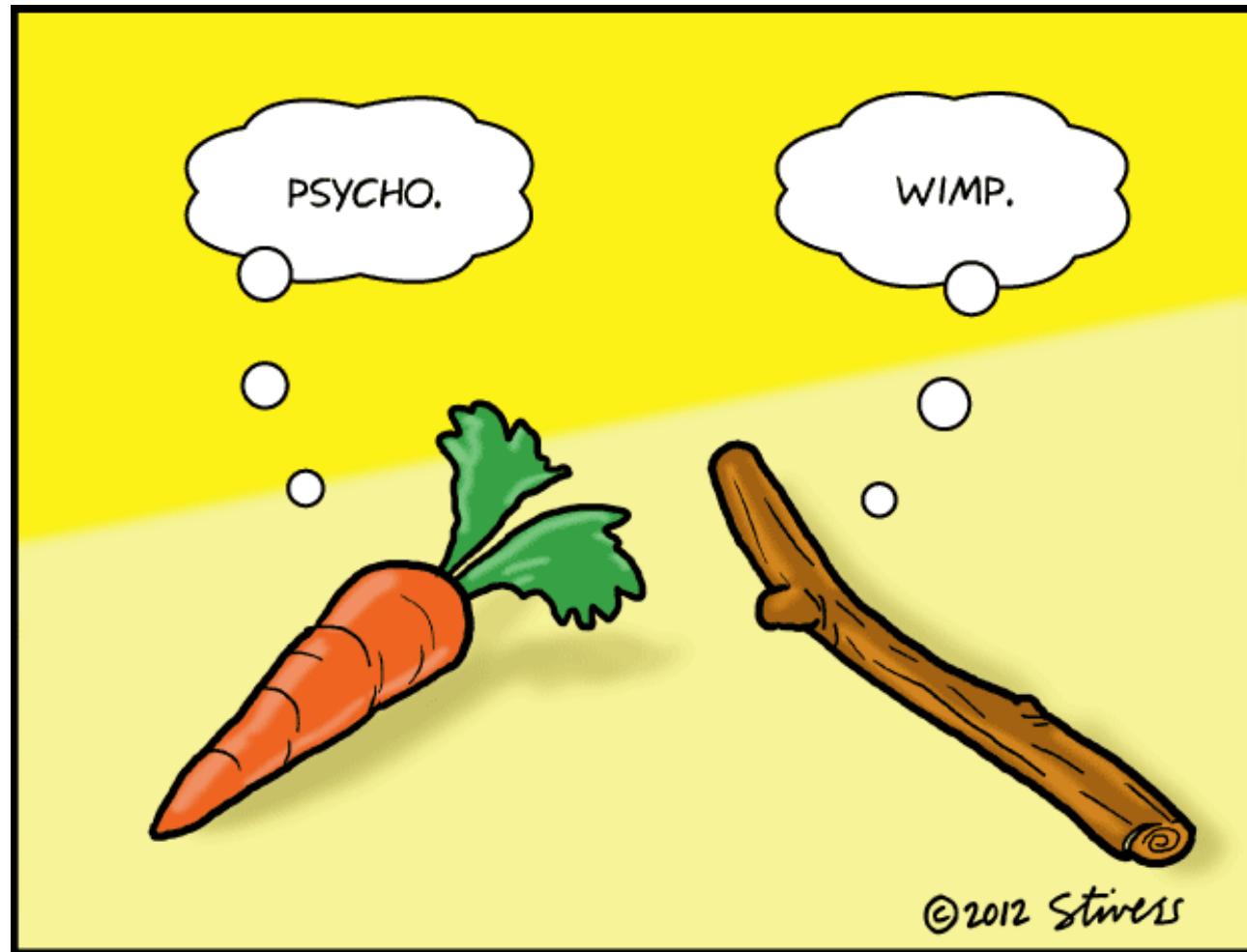
How Databrary is distinctive

- Open sharing among , not public
- Share identifiable data
- Store, search across, filter among participant & session characteristics
- Active (during study) curation reduces post hoc burden

What I've learned

Barriers to reproducibility

- Technological
- Cultural



"...psychologists tend to treat other peoples' theories like toothbrushes; no self-respecting individual wants to use anyone else's."

Walter Mischel, 2009

"Reviewers and editors want novel, interesting results. Why would I waste my time doing careful direct replications?"

Any number of researchers I've talked with

Tools empower

- [RStudio](#) & [Project Jupyter](#) and literate programming
- git, GitHub and version control

RStudio

File Edit Code View Plots Session Project Build Tools Help

Untitled1* | Source on Save | Run | Source | Values | History | Import Dataset |

```
1 rm(list = ls())
2 N <- 1000
3 u <- rnorm(N)
4 x1 <- -2 + rnorm(N)
5 x2 <- 1 + x1 + rnorm(N)
6 y <- 1 + x1 + x2 + u
7 r1 <- lm(y ~ x1 + x2)
8
```

10 |

10:1 (Top Level) R Script

Console ↵ ↻

```
Tapez <Entrée> pour voir le graphique suivant :
Tapez <Entrée> pour voir le graphique suivant :
Tapez <Entrée> pour voir le graphique suivant :
>
> ?lm
> rm(list = ls())
> N <- 1000
> u <- rnorm(N)
> x1 <- -2 + rnorm(N)
> x2 <- 1 + x1 + rnorm(N)
> y <- 1 + x1 + x2 + u
> r1 <- lm(y ~ x1 + x2)
>
```

Workspace | History | Values | Files | Plots | Packages | Help | R: Fitting Linear Models | Find in Topic | Im [stats] | R Documentation

N 1000
r1 lm[12]
u numeric[1000]
x1 numeric[1000]
x2 numeric[1000]
y numeric[1000]

Fitting Linear Models

Description

lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance (although [aoov](#) may provide a more convenient interface for these).

Usage

```
lm(formula, data, subset, weights,
method = "qr", model = TRUE, x =
singular.ok = TRUE, contrasts =
```

Arguments

```
## Joining multiple datasets
```

Fancy approach to multiple dataset merge. Joins datasets two at a time from left to right in the list. The result of a two-table join becomes the 'x' data

```
```{r data-frame-demo}
df1 <- data.frame(id=1:10, x=rnorm(10), y=runif(10))
df2 <- data.frame(id=1:11, z=rnorm(11), a=runif(11))
df3 <- data.frame(id=2:10, b=rnorm(9), c=runif(9))
```

```
Reduce(function(...) { full_join(...) }, list(df1, df2, df3))
```

```
```
```

R Markdown

- One document format
 - Text, images, movies, data plots, code (not just R), commentary, citations, equations
- Many outputs
 - HTML slides (like this one)
 - PDF, MS Word, Markdown documents, even full manuscripts!
 - Web sites, blogs
 - Books

jupyter spectrogram (autosaved)



File Edit View Insert Cell Kernel Help | Python 3



Simple spectral analysis

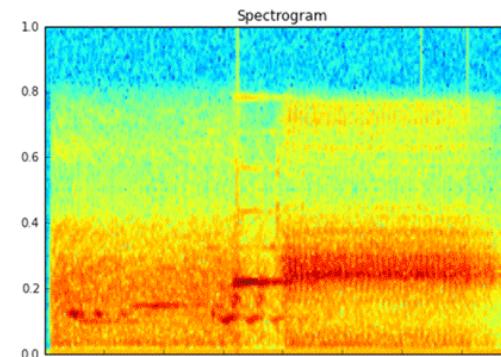
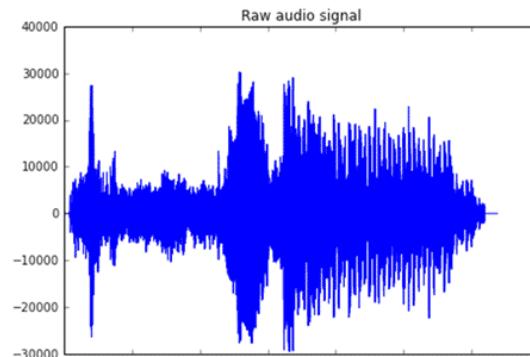
An illustration of the [Discrete Fourier Transform](#)

$$X_k = \sum_{n=0}^{N-1} x_n \exp^{-\frac{2\pi i}{N} kn} \quad k = 0, \dots, N-1$$

```
In [2]: from scipy.io import wavfile  
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view its spectral structure using matplotlib's builtin specgram routine:

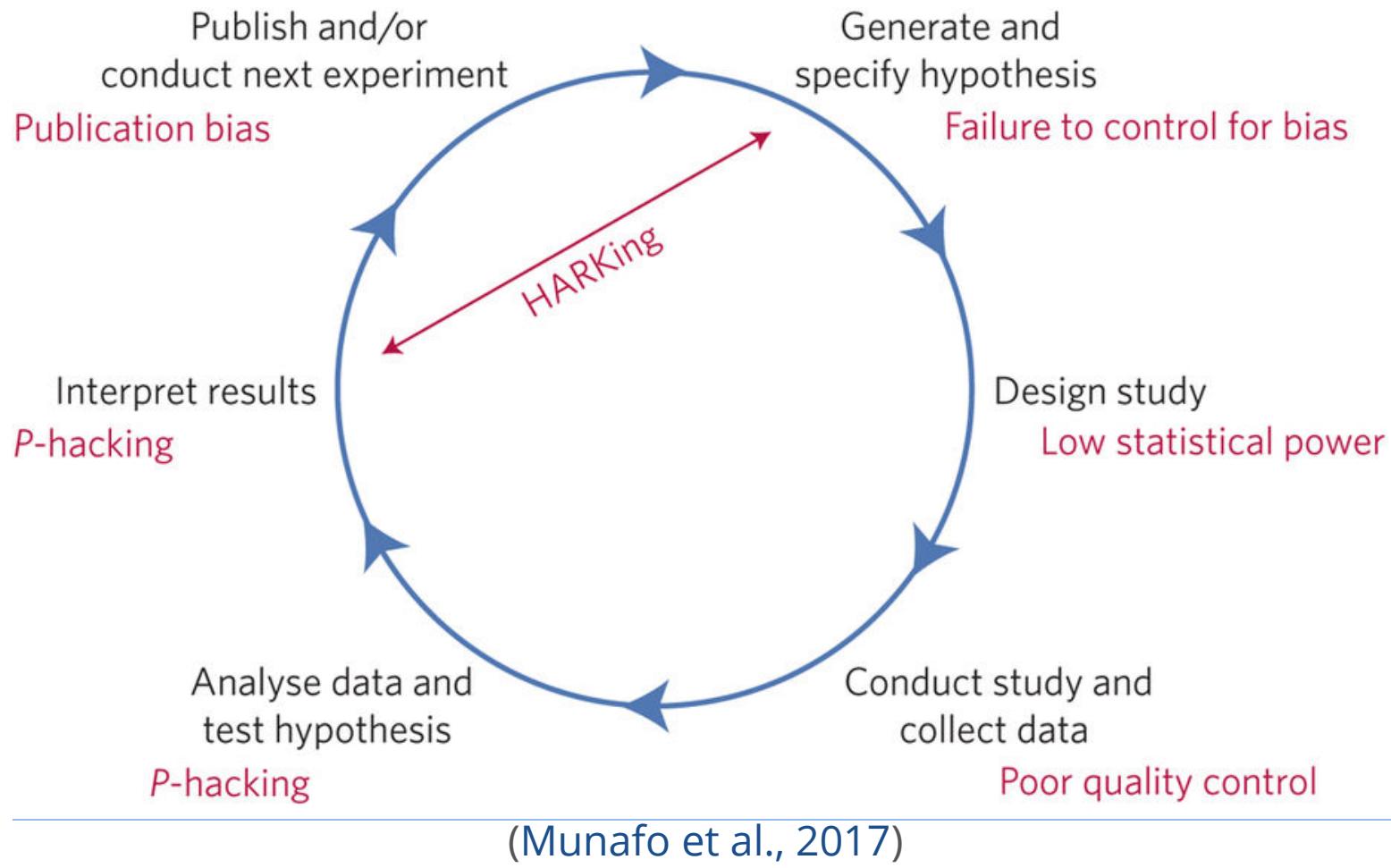
```
In [5]: fig, (ax1, ax2) = plt.subplots(1,2,figsize(16,5))  
ax1.plot(x); ax1.set_title('Raw audio signal')  
ax2.specgram(x); ax2.set_title('Spectrogram');
```



Next generation of scientific publishing

- Lab notebooks that embody literate programming principles
- Close links between data collection, cleaning, analysis, data repositories, preprints, publishers
- Persistent identifiers for research materials, code, & resources
 - All published figures, data tables, data sets, analysis code...

Let's not waste a "good" crisis







Collect & share video as data and documentation

PUBLISHED: 12 JUNE 2017 | VOLUME: 1 | ARTICLE NUMBER: 0128

comment

Video can make behavioural science more reproducible

Rick O. Gilmore and Karen E. Adolph

We recommend the widespread use of a simple, inexpensive, easy-to-implement, and uniquely powerful tool to improve the transparency and reproducibility of behavioural research — video recordings.

How robust are published findings in the behavioural sciences? Can the results be reproduced based solely on information in the published paper? We suggest that questions about scientific reproducibility¹ — especially in the behavioural sciences — often arise because text-based descriptions and static images in the methods sections of journal articles cannot specify procedures and dynamic stimuli in sufficient detail for the study to be readily and reliably replicated by researchers



Increase sample sizes

- Collect larger samples



Standardize metadata

- participants (age, gender, race/ethnicity, ...)
- settings (times, dates, places)
- measures & tasks



Improve statistical practices

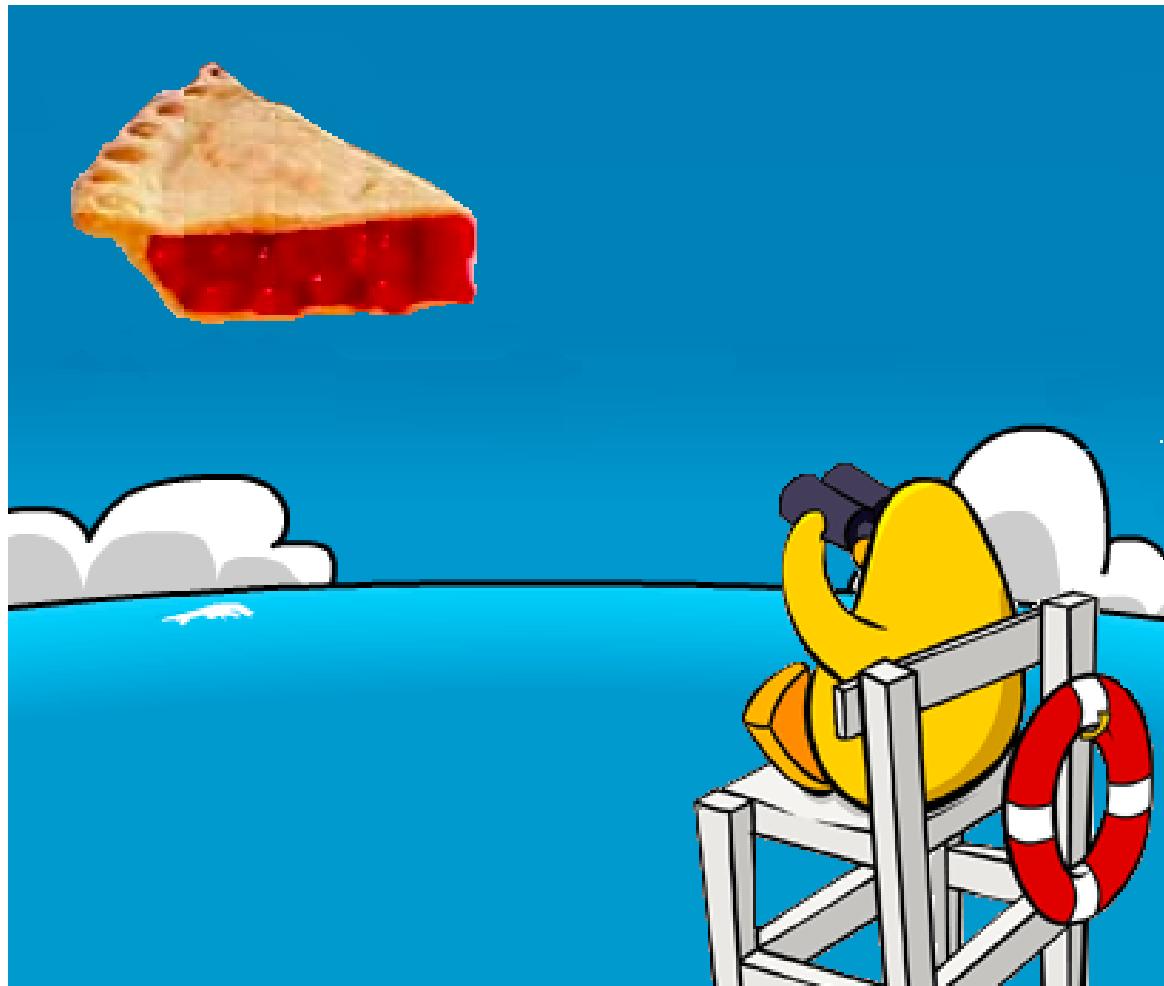
- Automated checking of paper statistics (in American Psychological Association formats) via [Statcheck](#)
- Redefine "statistical significance" as $p < .005$? [\(Benjamin et al., 2017\)](#)
- Or move away from NHST toward more robust and cumulative practices

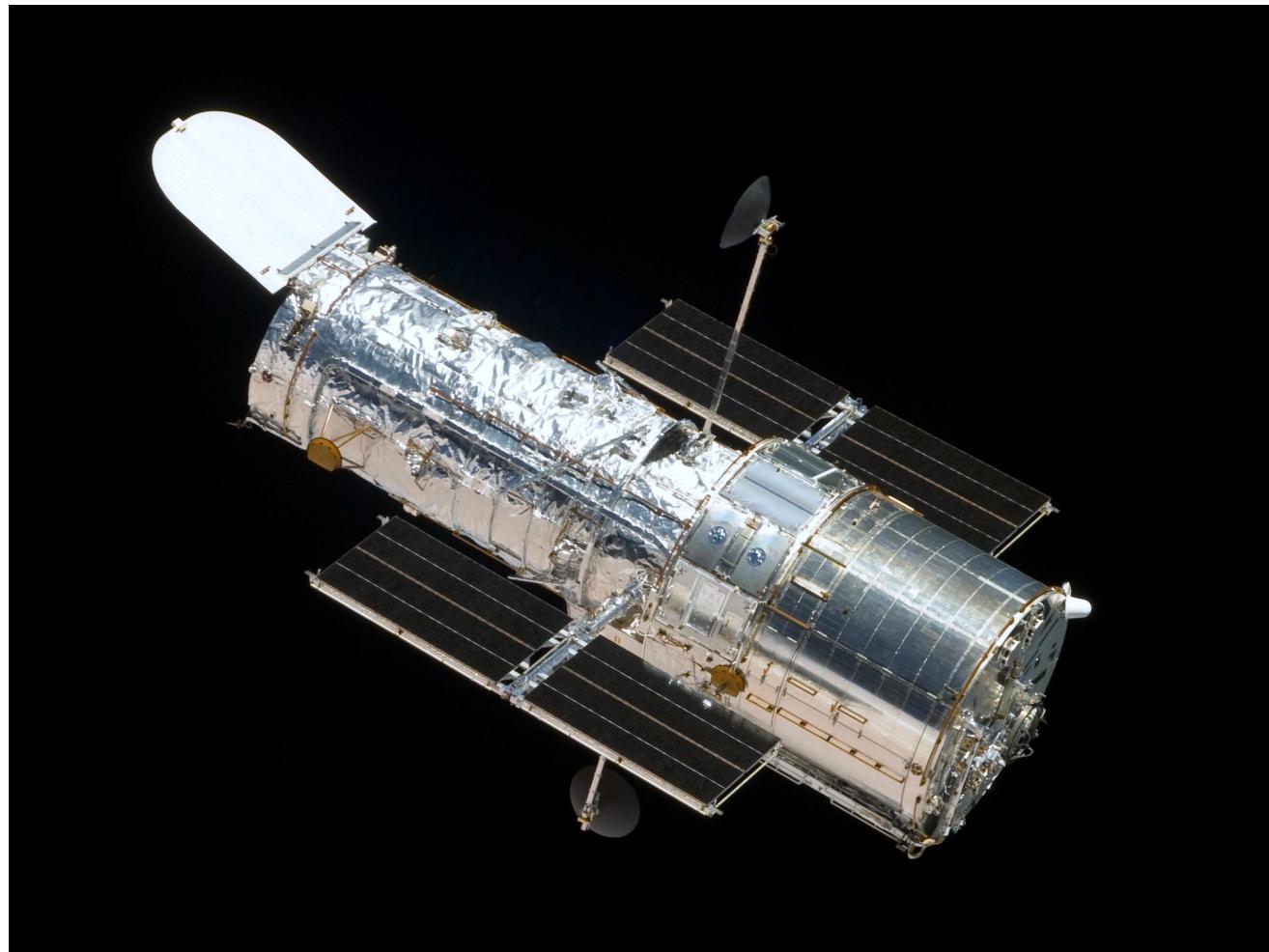
Store data, materials, code in repositories

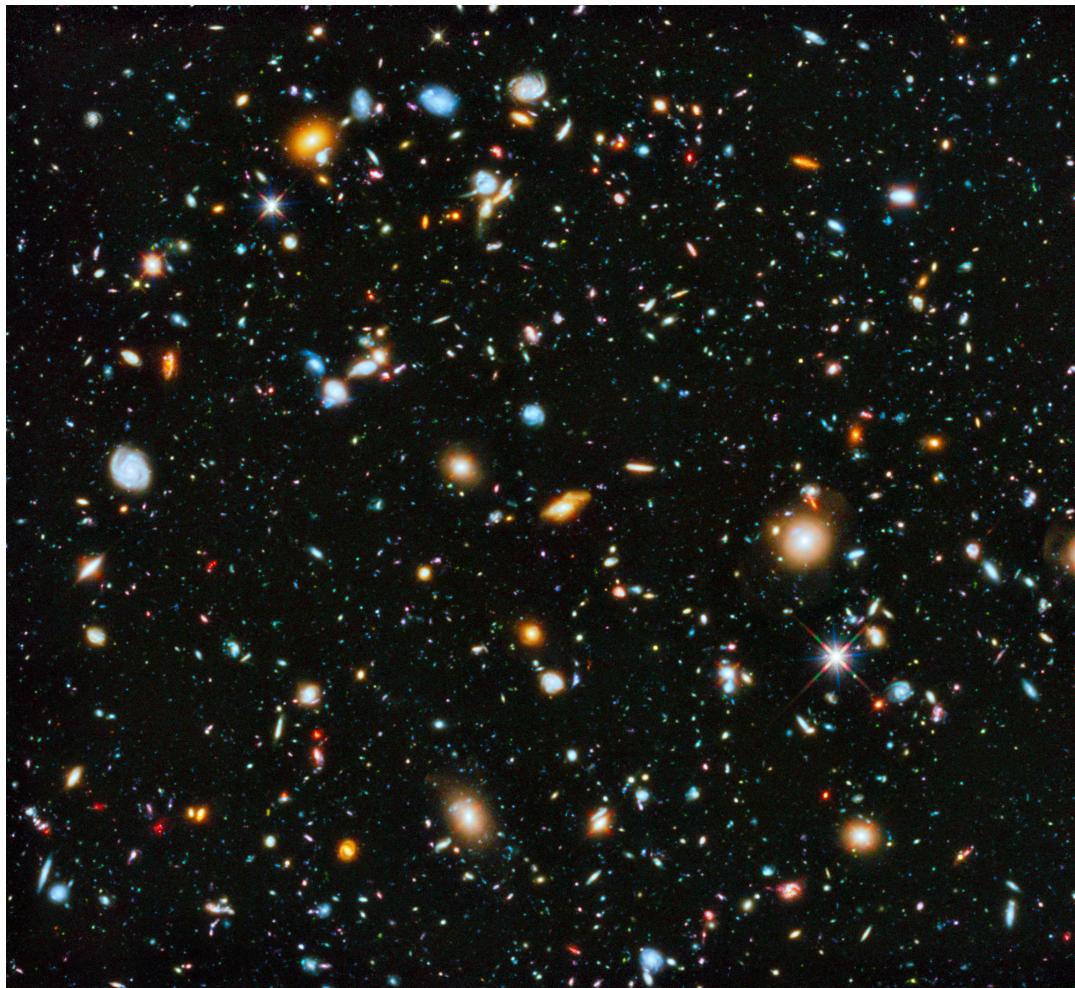
- Funder, journal mandates for sharing increasing
- But no long-term, stable, funding sources for curation, archiving, sharing
- ArXiv model
 - Institutional (Cornell) support
 - Subscription

Build platforms for discovery

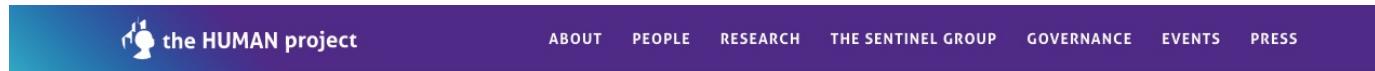
- Data + analysis
- e.g., PSU's [Biostars](#)







Data from diverse domains



10,000

New Yorkers

4,000

Households

20

Years

1

Research Platform





Link measures across people



Contribute to research & citizen science!

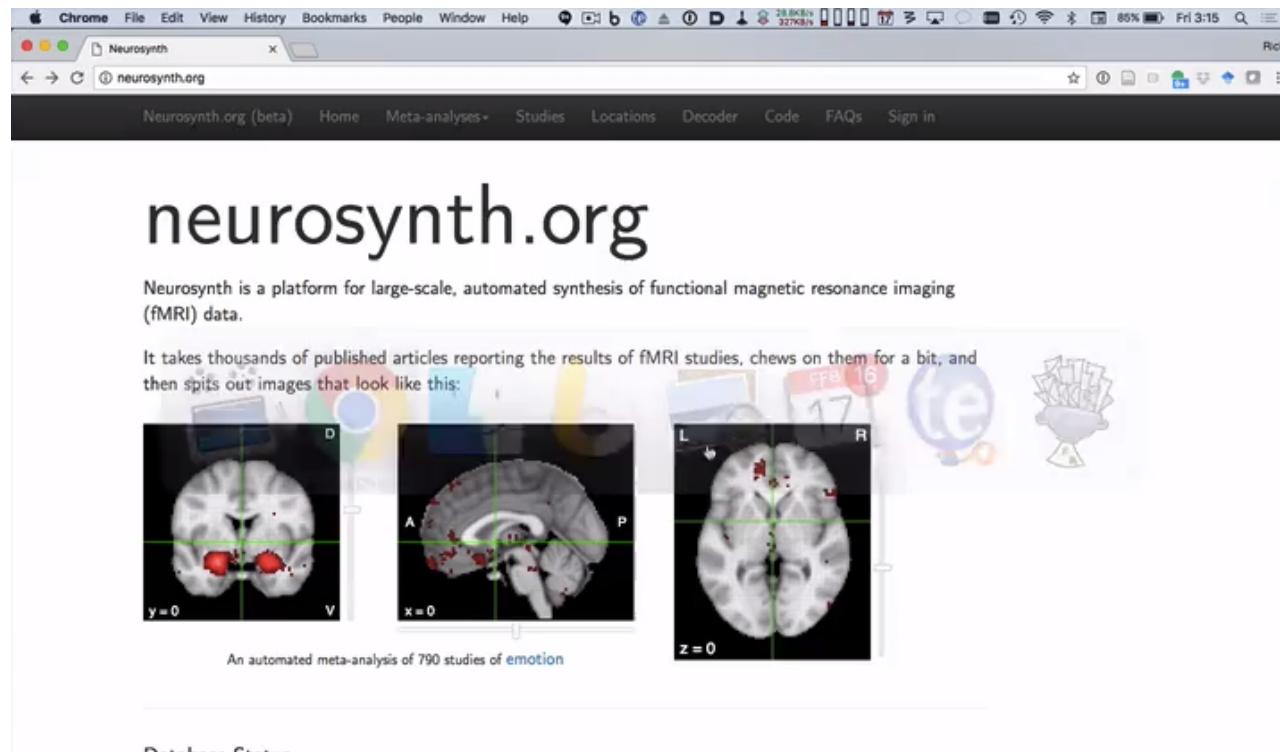
Through Open Humans, you can gather valuable data about yourself and find cool projects to share it with.

1. Sign up

2. Connect data

3. Join projects

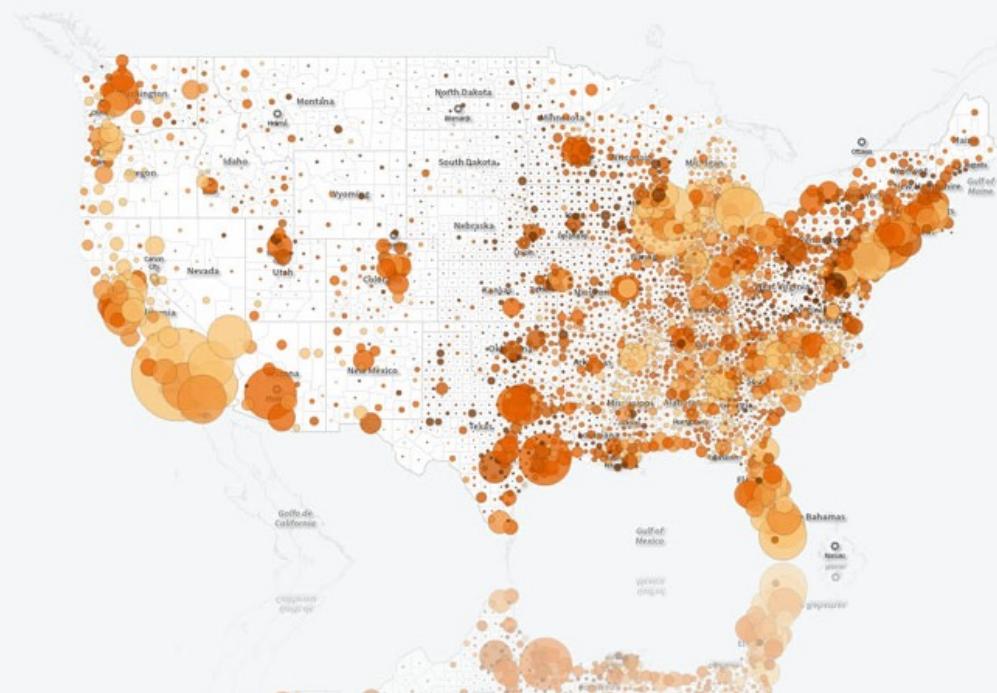
Web-based data visualization, analysis





Visually Explore Demographic Data

220 years of demographic data, 25,000 maps, hundreds of profile reports, 40 billion data elements and 335,000 variables

[Take the Tour](#)[Start Now →](#)

Wordbank: An open database · X

wordbank.stanford.edu

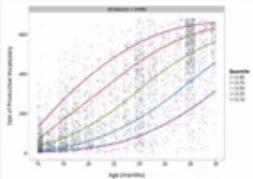
Wordbank Analyses Contributors Publications Population Blog FAQ



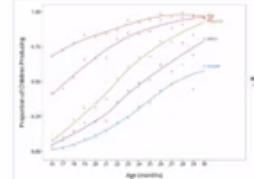
Wordbank

An open database of children's vocabulary development



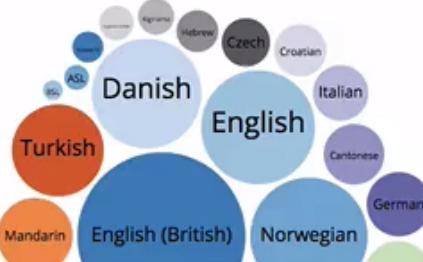


Vocabulary Norms
Explore vocabulary size growth curves for various languages and demographic groups.



Item Trajectories
Explore trajectories of individual words, word categories, and grammar items.

Wordbank contains data from 63,386 children and 71,003 CDI administrations, across 23 languages and 44 instruments:



Search, filtering by personal characteristics

NIMH Data Archive

Home Query Harmonization Tools Cloud Contribute Request Access Policy Tutorials About FAQ Tools ▾

Quick Navigation

- Query**
 - General Query
 - Data from Labs
 - Data from Papers
- By Measure/Element**
- By Concept**
- By GUID**
- omicSearch
- Share**
 - Harmonization Standards
 - Prepare and Submit
- Compute**
- Resources**
 - GUID Tool
 - Validation and Upload Tool
 - Download Manager
 - Data Dictionary
- Contact Us
- Request Account

The National Database for Autism Research (NDAR) is an NIH-funded research data repository that aims to accelerate progress in autism spectrum disorders (ASD) research through data sharing, data harmonization, and the reporting of research results. NDAR also serves as a scientific community platform and portal to multiple other research repositories, allowing for aggregation and secondary analysis of data.

Data Distribution
151,831 subjects by age, 95,929 individuals

Gender¹

| Gender | Count |
|--------------|--------|
| Female | 44,812 |
| Male | 78,412 |
| Not Reported | 28,607 |

Phenotypic¹

| Category | Count |
|-----------------------------------|--------|
| Autism Spectrum Severely Affected | 31,885 |
| Autism Spectrum Affected | 15,082 |
| Autism Spectrum Mildly Affected | 23,238 |
| NDAR Controls | 63,363 |
| Fragile X | 11,520 |
| RDoC Subjects | 1 |
| ABCD Subjects | 1 |
| Not Defined | 1 |

NDAR Controls: 63,363 (41.7%)

Neuroimaging

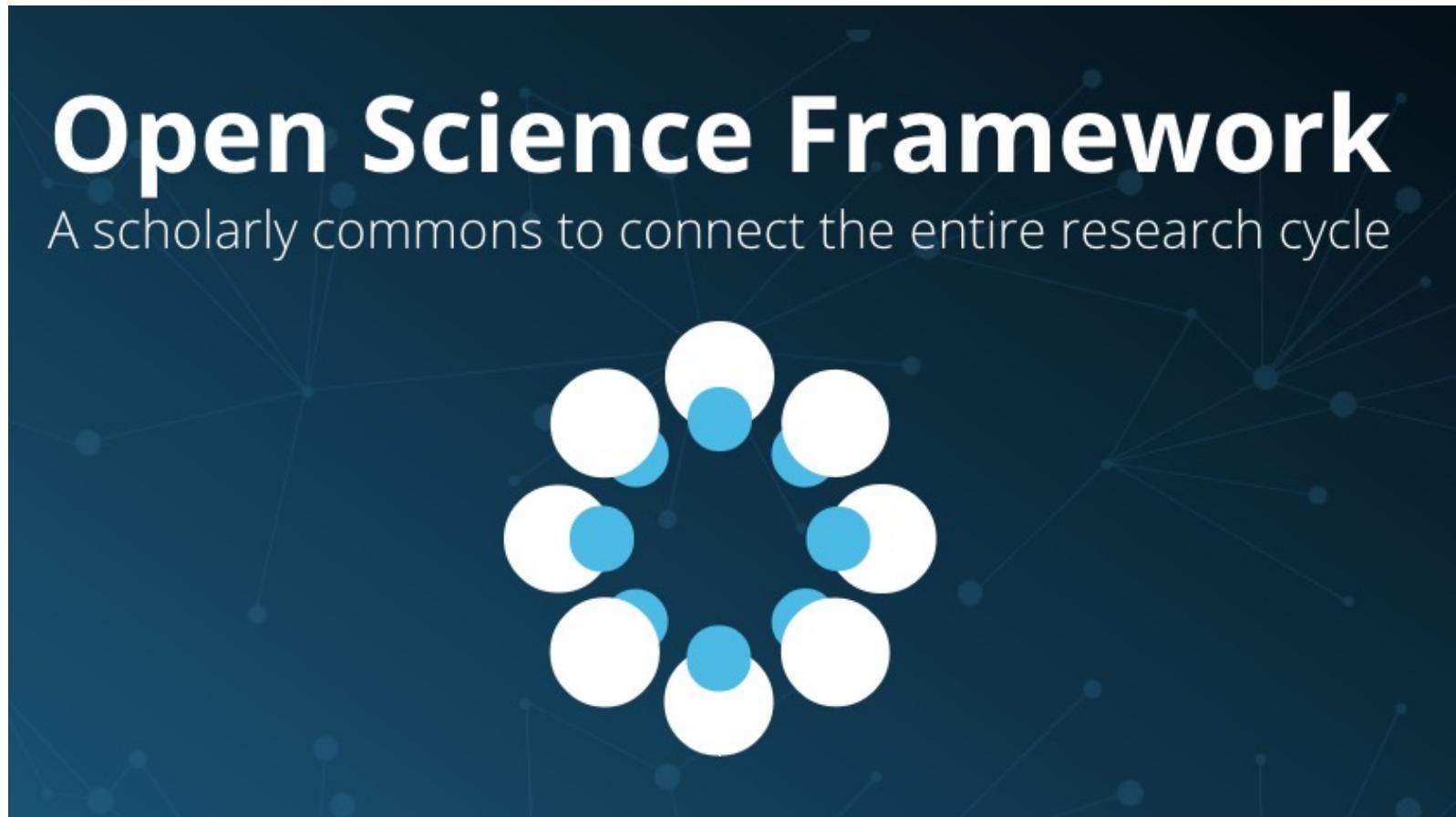
| Method | Count |
|--------------|-------|
| DTI | 1,753 |
| fMRI | 1,480 |
| MRI | 9,389 |
| Spectroscopy | 1,716 |
| EEG | 1 |
| MEG | 1 |
| PET | 1 |
| Eye Tracking | 1 |

Genomic

| Technique | Count |
|--|--------|
| SNP and CNV: microarray | 10,553 |
| Next Generation Sequencing: sequencing | 13,324 |
| Gene regulation: microarray | 1 |
| Gene expression: sequencing | 1 |
| SNP and CNV: sequencing | 1 |
| Gene regulation: sequencing | 1 |
| Gene expression: microarray | 1 |
| STR Genotyping: PCR | 1 |
| Sanger sequencing: sequencing | 1 |

¹ Numbers reported are subjects by age

Self/active curation of data, materials



Consistent, clear sharing permissions structure



Databrary is a video data library for developmental science.

Share videos, audio files, and related metadata. Discover more, faster.

593 investigators 248 affiliates 332 institutions

401 total volumes 75 shared volumes 43,193 files 8,151 hours of recordings 8,595 participants



Your Profile

Manage your data and authorizations.



Browse Data



Browse People

Progress

| Example | Multi-measure | Indiv link/search | Visualize | Self-curate | Permissions |
|-------------|---------------|-------------------|-----------|-------------|-------------|
| Databrary | ✓ | ✓ | tabular | ✓ | ✓ |
| Human Proj | ✓ | ✓ | ? | ? | ✓ |
| ICPSR | ✓ | ? | ✓ | ? | ✓ |
| Neurosynth | fMRI BOLD | group data | ✓ | public | NA |
| OpenNeuro | ✓ | ? | ✓ | ✓ | public |
| Open Humans | ✓ | ✓ | ? | ? | ✓ |
| OSF | ✓ | | | ✓ | public |
| WordBank | M-CDI | group metadata | ✓ | ? | public |



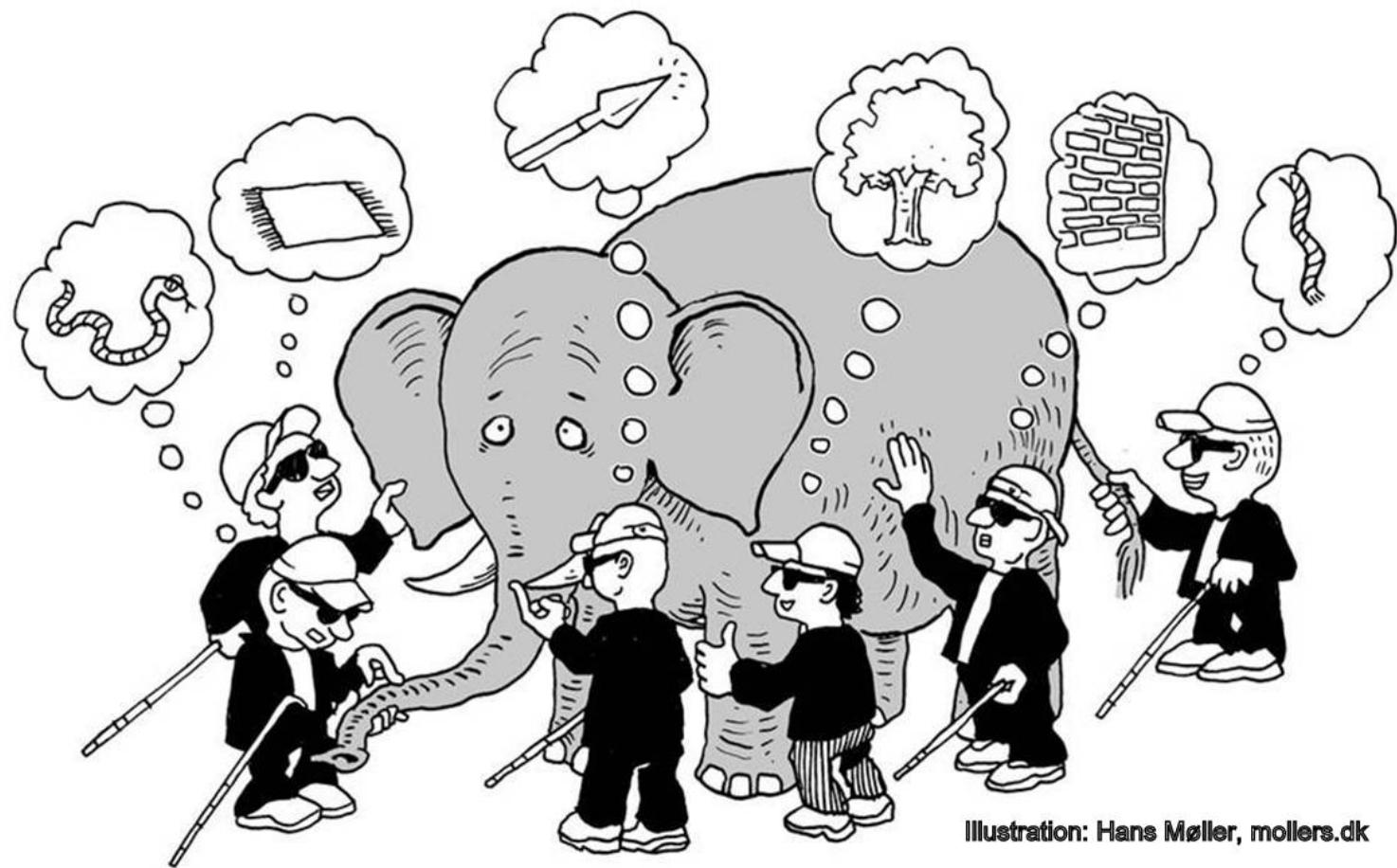
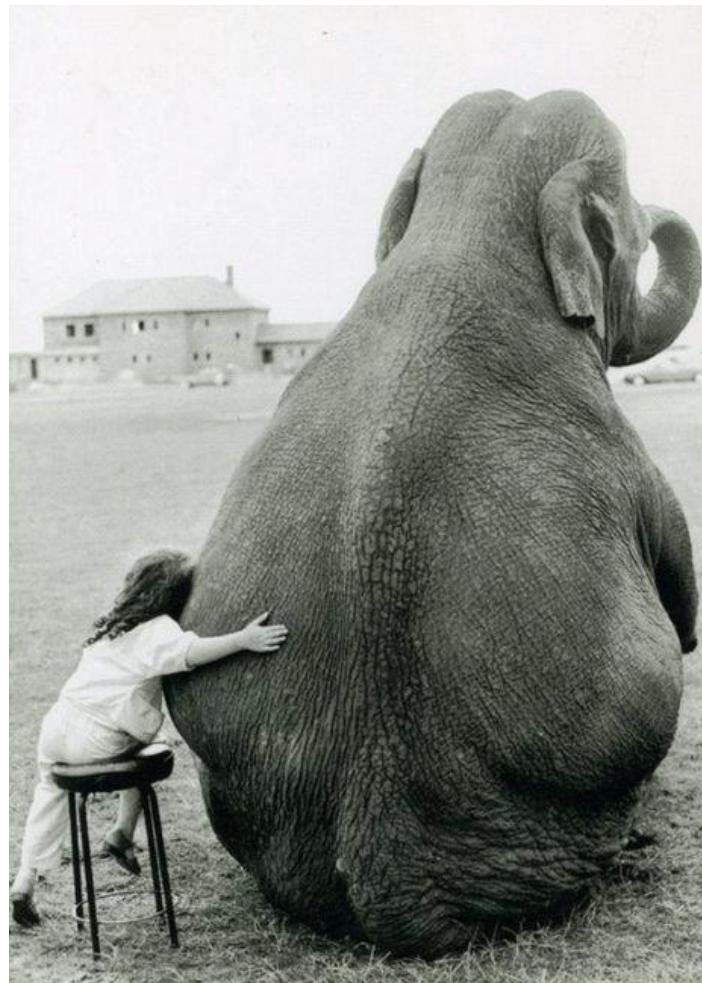


Illustration: Hans Møller, mollers.dk





Keep in touch

rogilmore@psu.edu

gilmore-lab.github.io

Stack

This talk was produced on 2017-09-07 in [RStudio Server Pro](#) using R Markdown and the reveal.JS framework on Penn State's [ACI-ICS RStudio Server Pro instance](#). The code and materials used to generate the slides may be found at <https://github.com/gilmore-lab/aci-ics-2017-09-07/>. Information about the R Session that produced the code is as follows:

```
## R version 3.4.1 (2017-06-30)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions/A/libBLAS.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats   graphics grDevices utils   datasets methods  base
##
## other attached packages:
## [1] DiagrammeR_0.9.2 revealjs_0.9
##
```