

Bad Data, Big Models & Statistical Methods for Studying Evolution

Richard McElreath

*Max Planck Institute for Evolutionary Anthropology
Leipzig*



DAVID BARLOW ARCHIVE

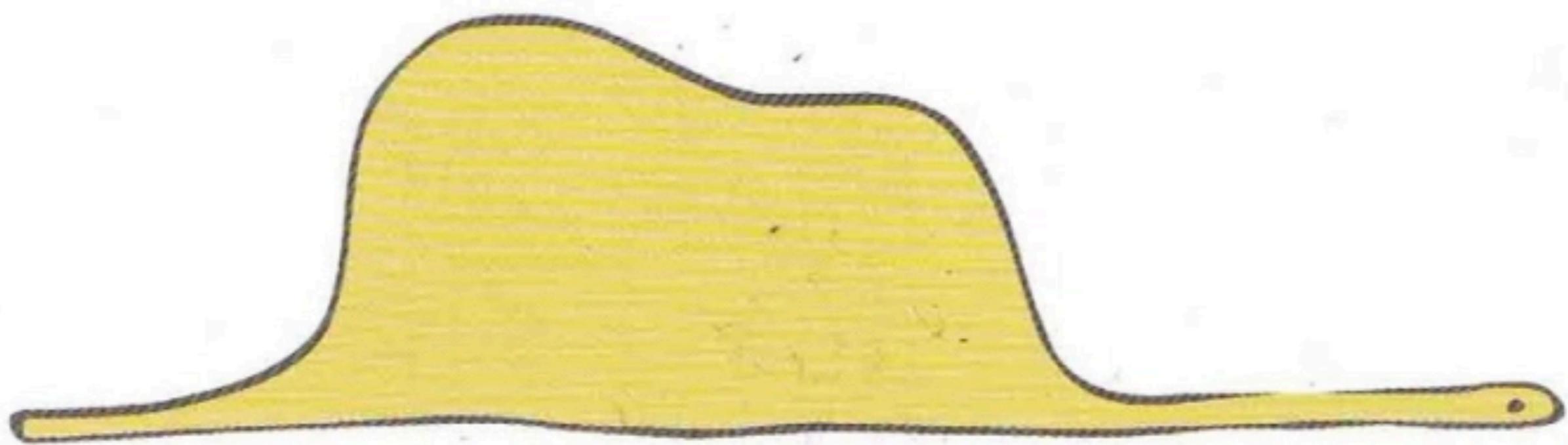
015010-MA09C009

Source TC: 09:02:46:07

Elephant embryo, 16 weeks

Sturnus vulgaris





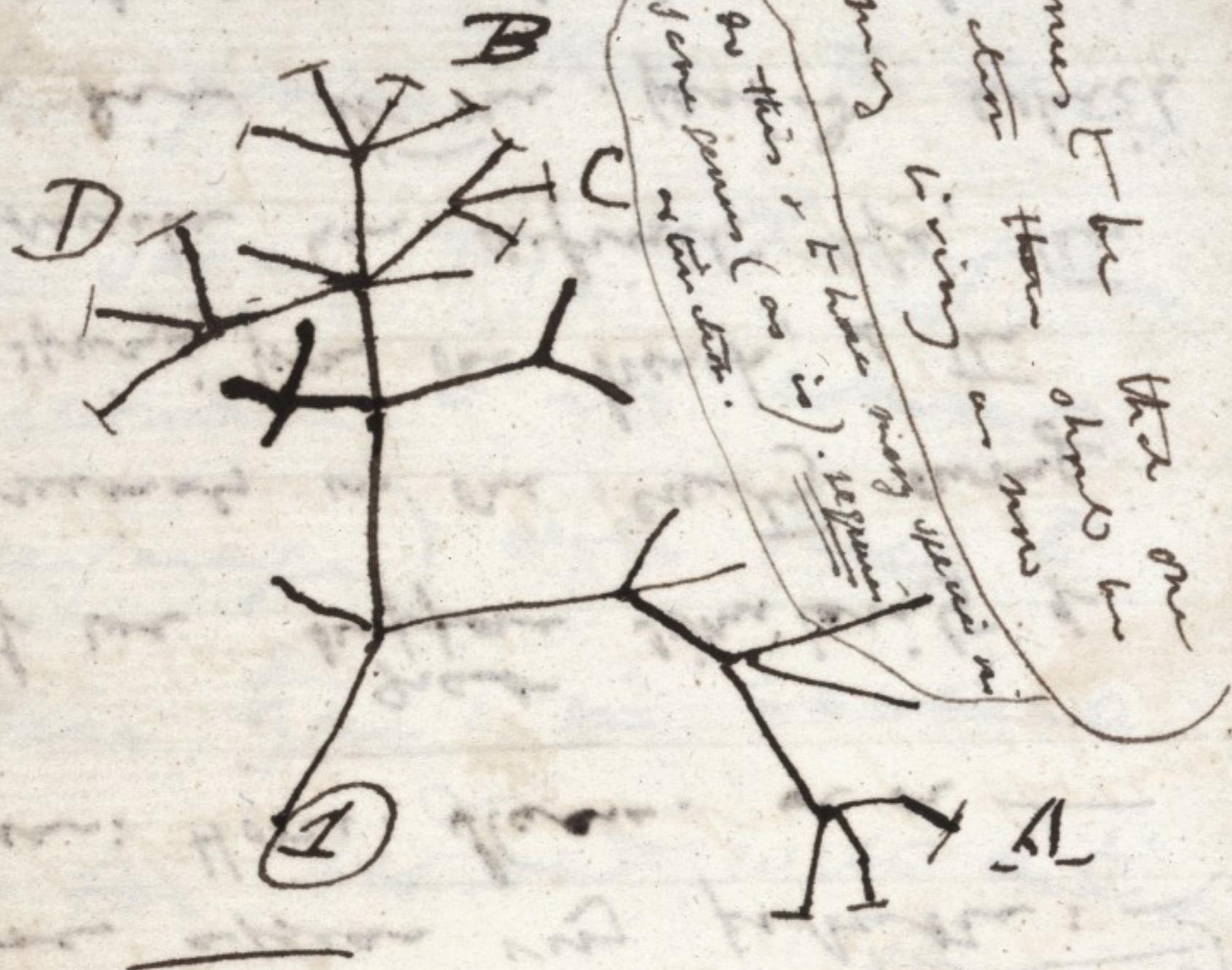
*Mon dessin ne représentait pas un chapeau. Il représentait
un serpent boa qui digérait un éléphant*



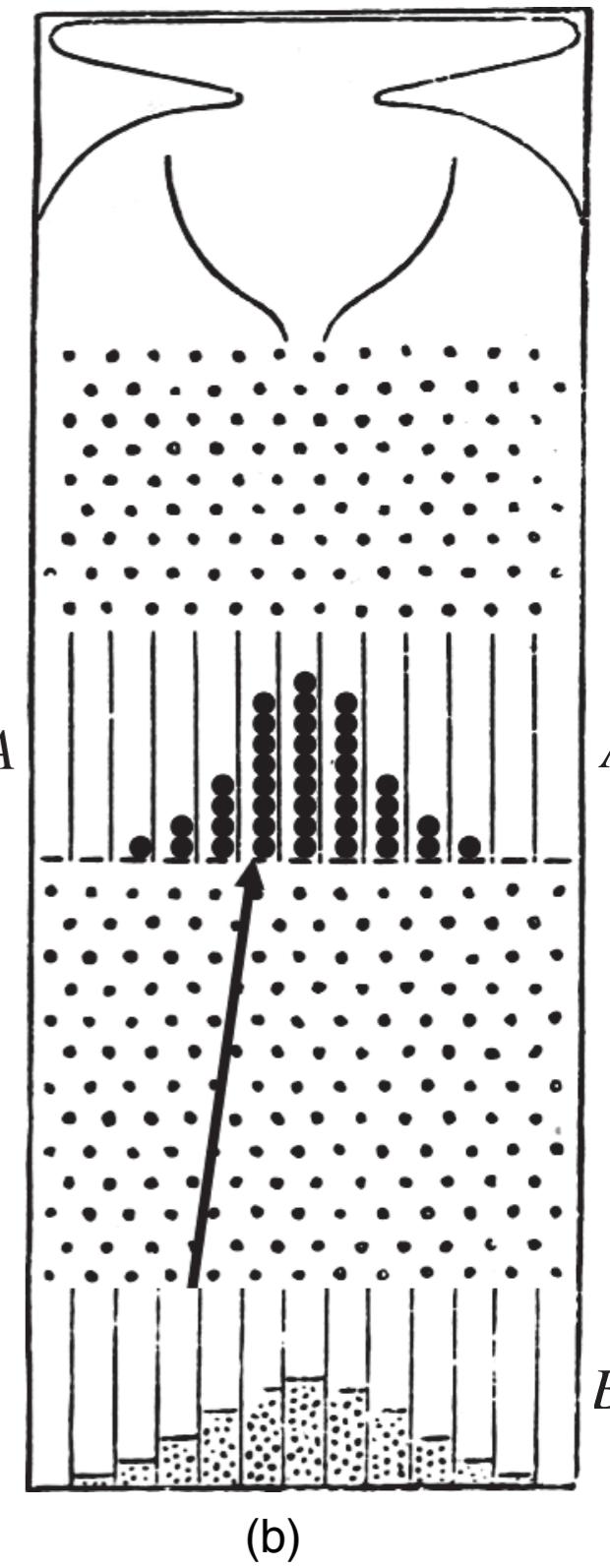
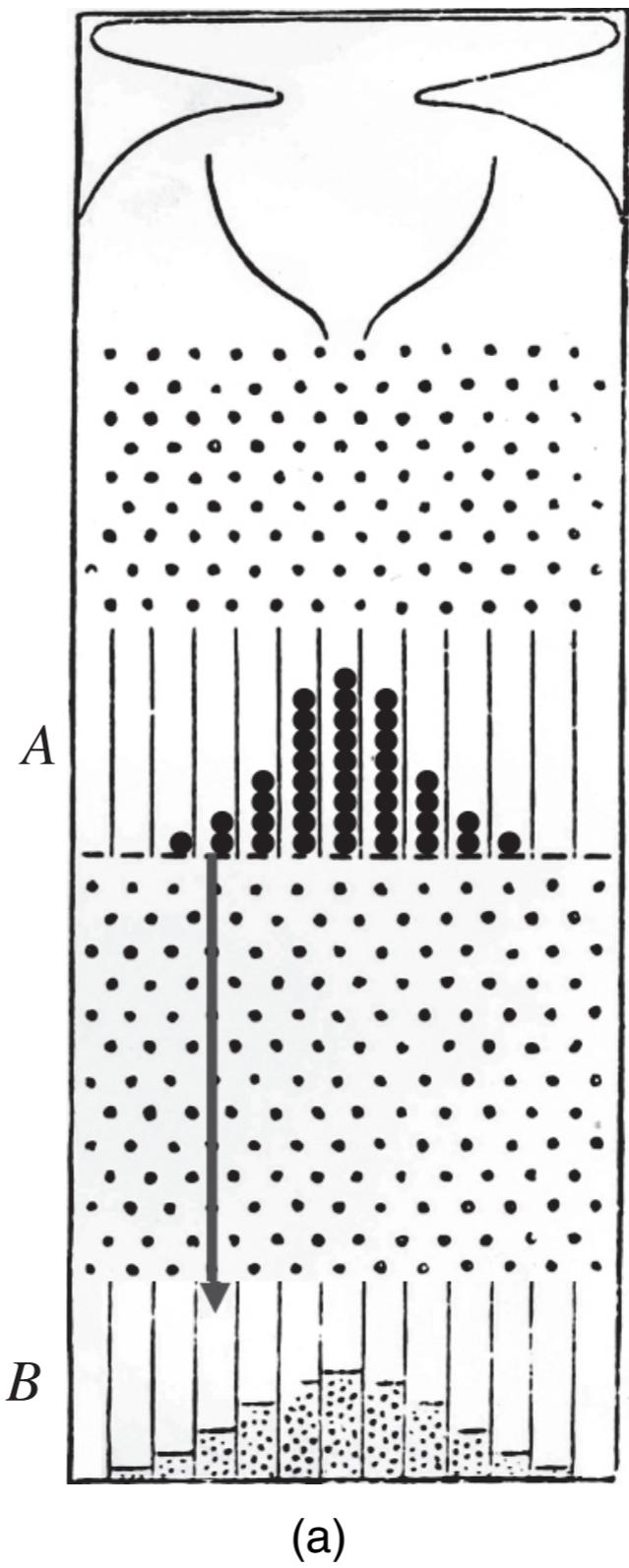
Homo sapiens



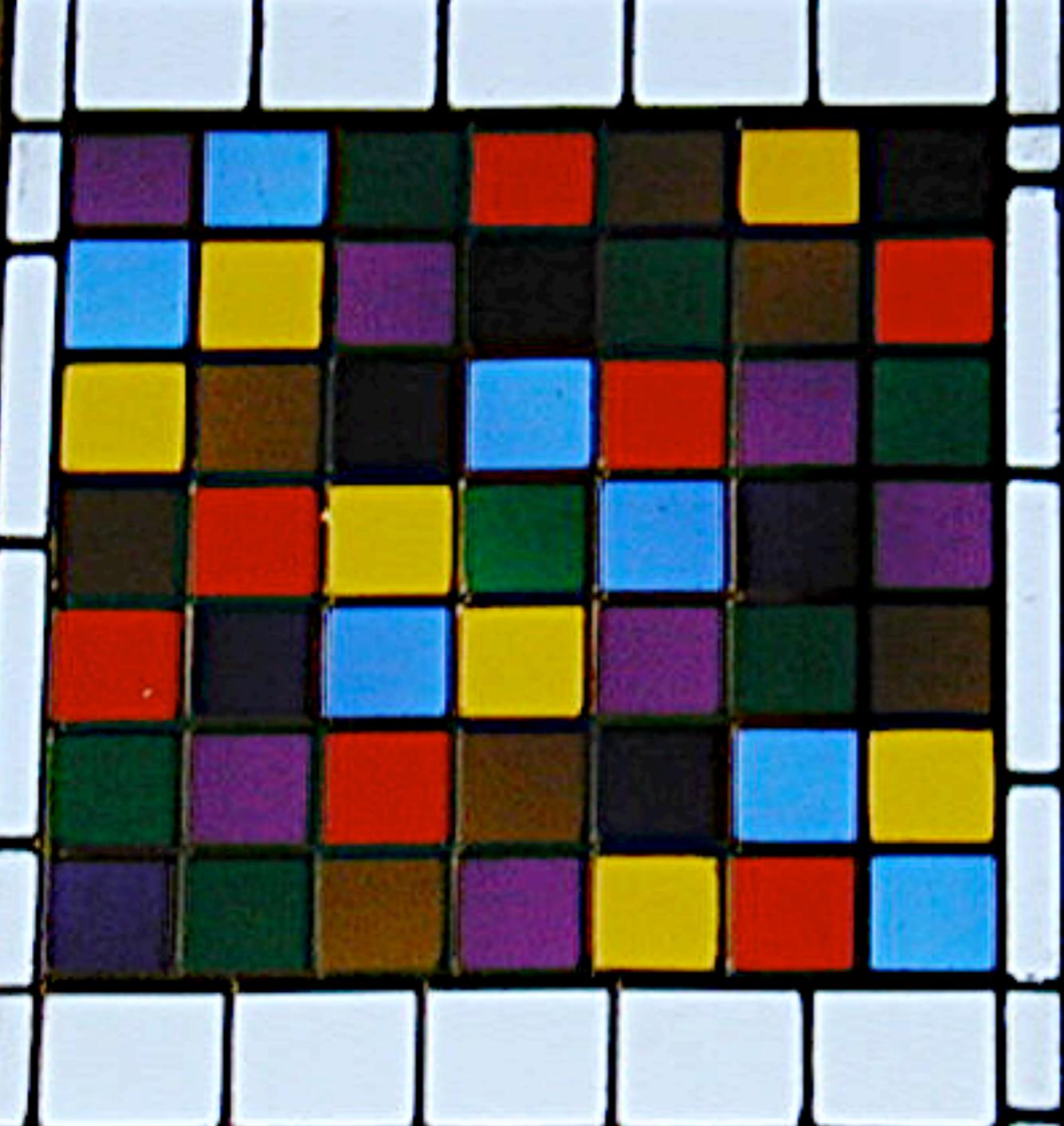
I think



thus between A & B. various



Francis Galton builds Bayesian conditional distributions (1873, 1889).
See Stigler (2010): Darwin, Galton and the Statistical Enlightenment



R.A.FISHER
FELLOW 1920-26 1943-62
PRESIDENT 1956-59



Rothamsted Experimental Station

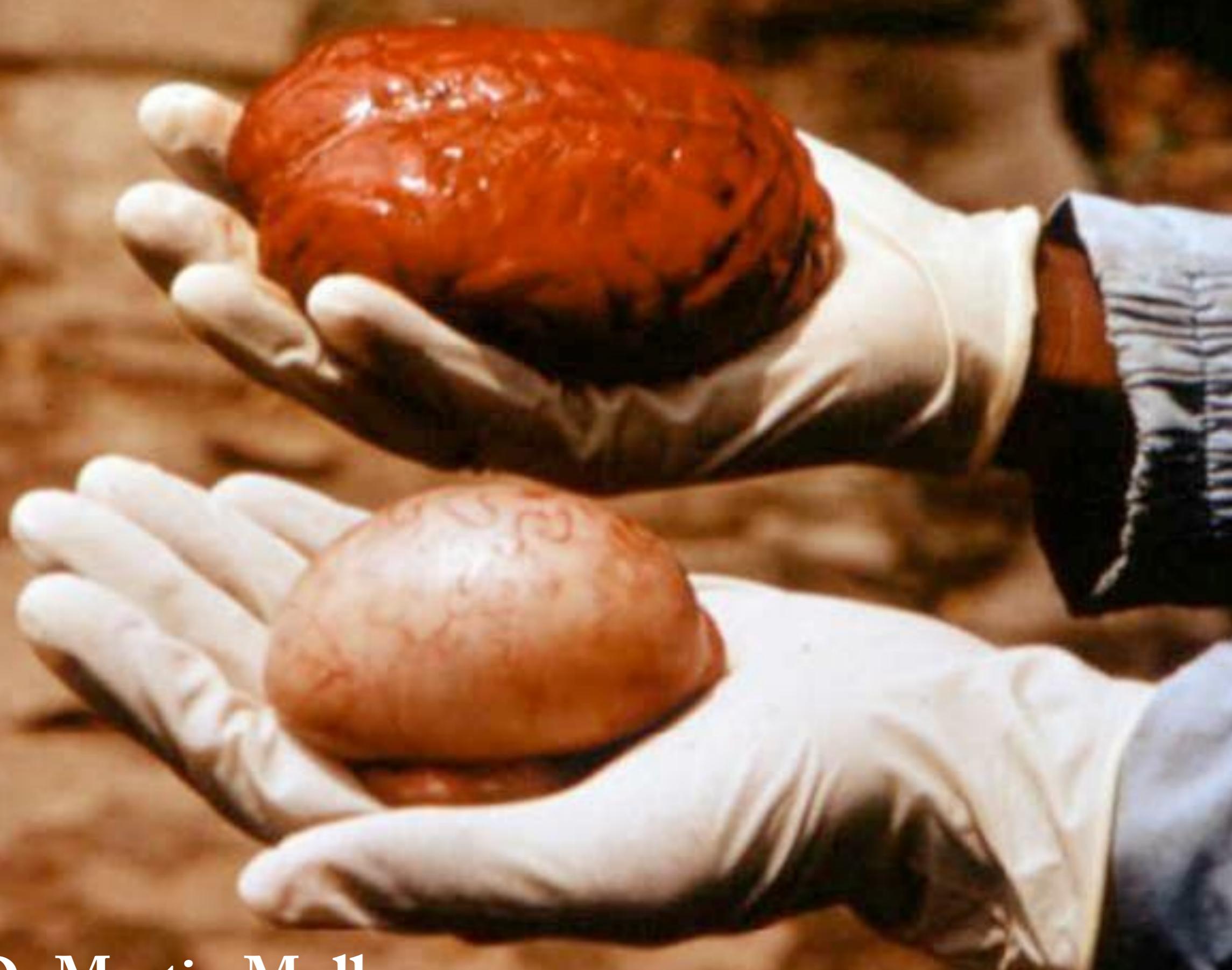
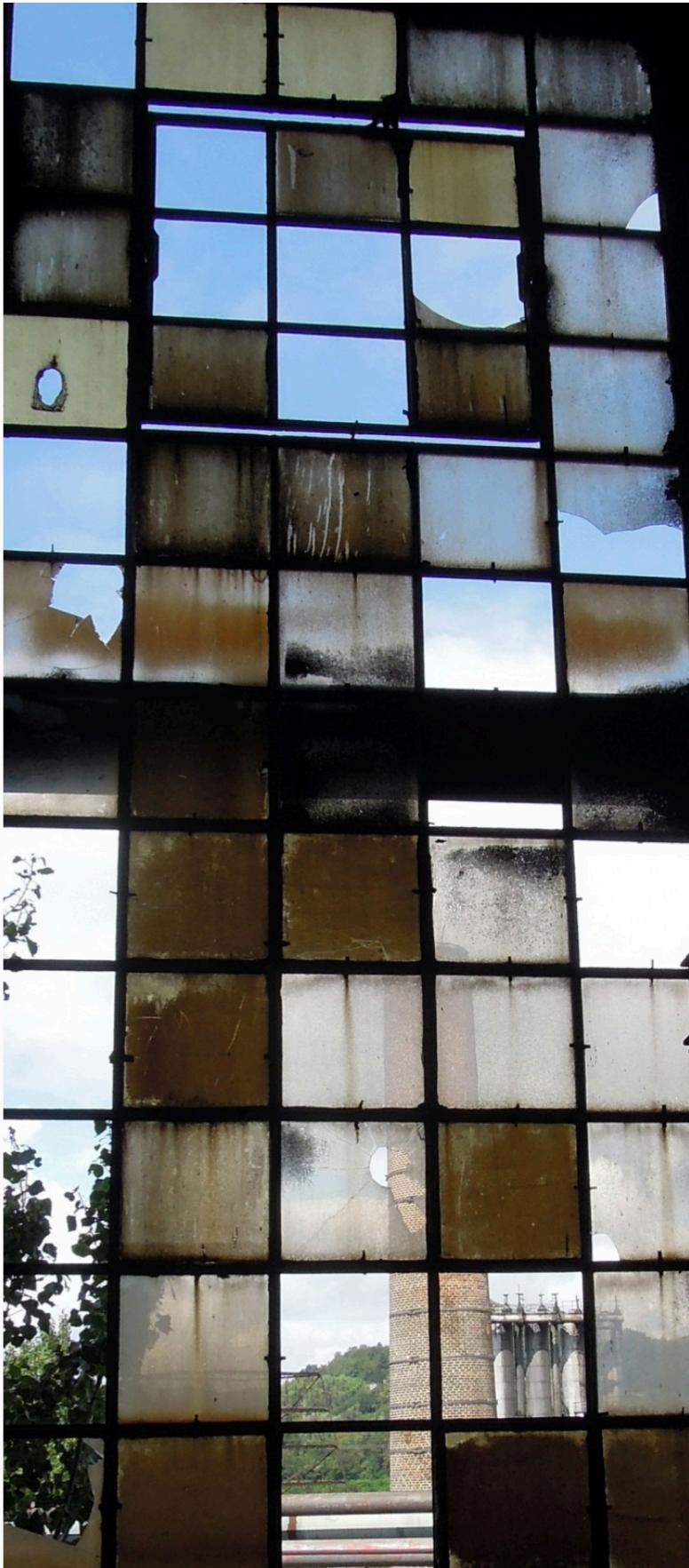


Photo: Dr Martin Muller

Statistical Inference in Evolution & Ecology

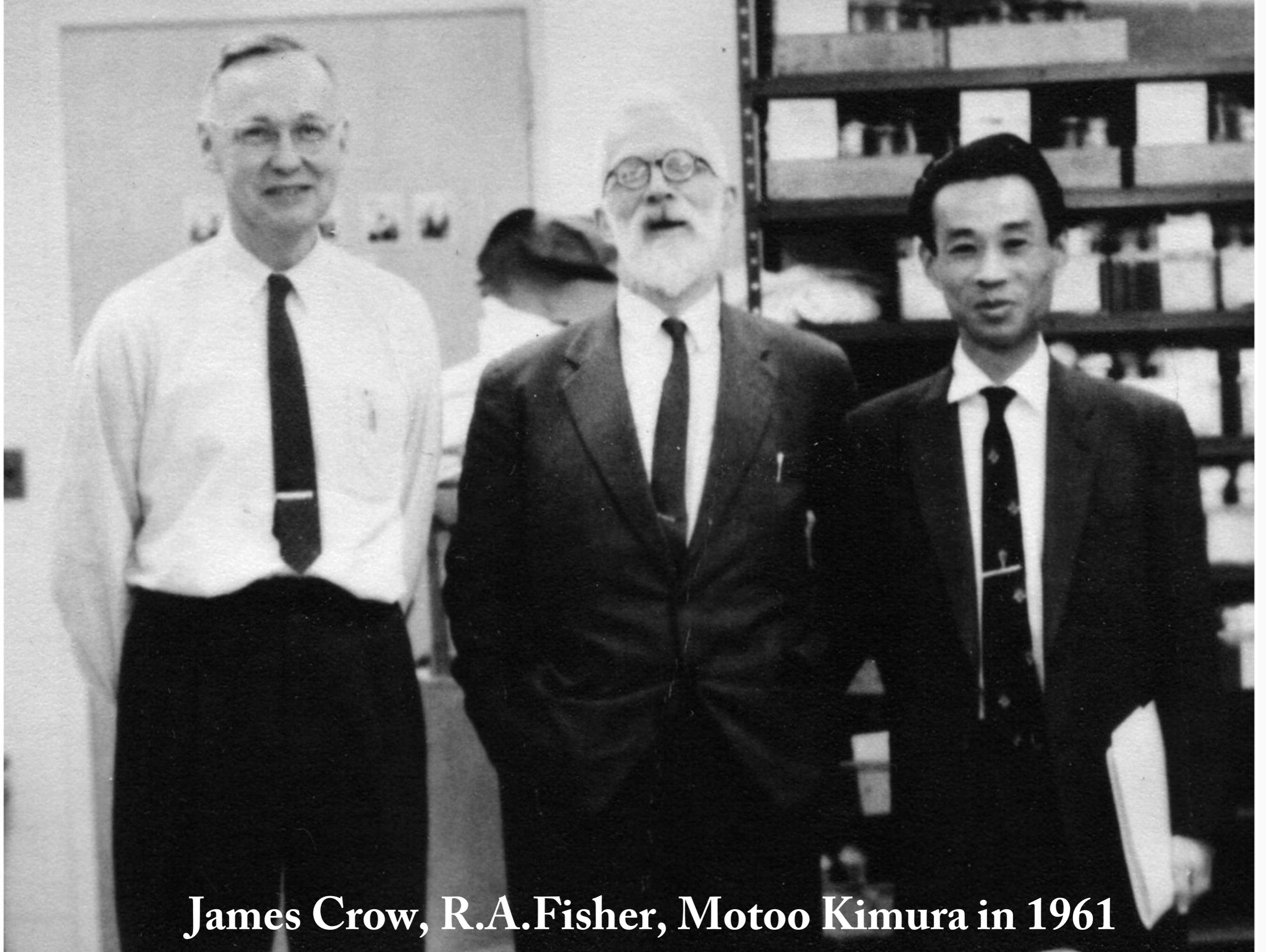
- Fundamentally *observational*
- Data: *Bad, scarce & derived*
- Null models not unique
- Plausible models *custom & hungry*
- Result: Hi-D headaches, even with small data
- Need: Flexible, scalable tools



Outline

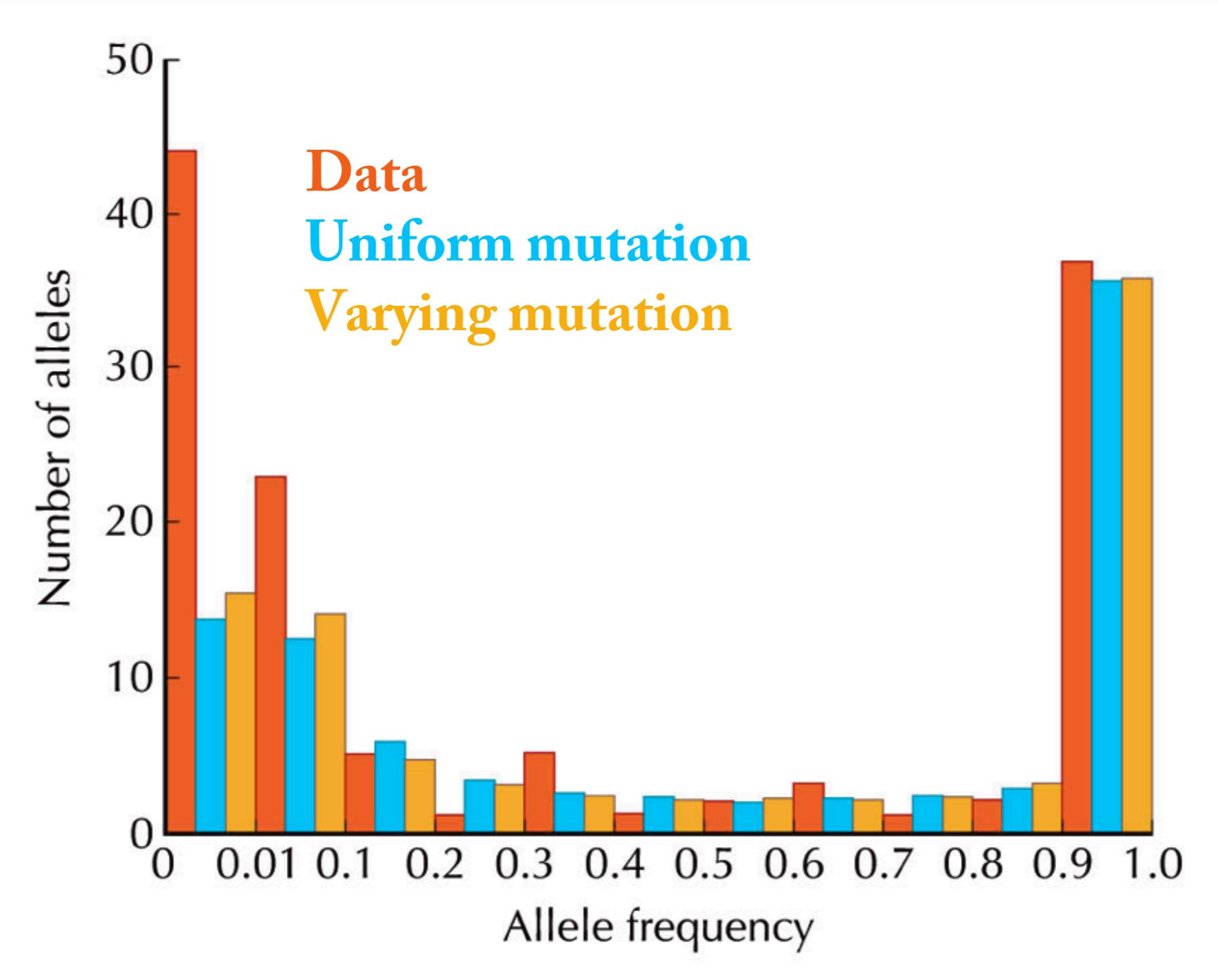
- The evolution of statistical methods for studying evolution
- Bad data in a big Stan model





James Crow, R.A.Fisher, Motoo Kimura in 1961

Neutral theory



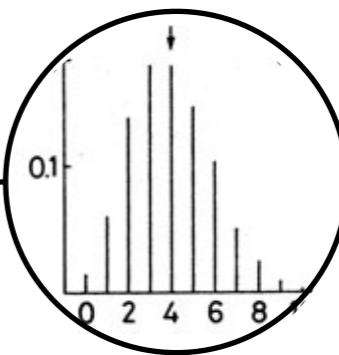
Hypotheses

$X \rightarrow Y, Z$

Process models

$$\begin{aligned}\frac{dx}{dt} &= a(y - x) \\ \frac{dy}{dt} &= x(b - z) - y \\ \frac{dz}{dt} &= xy - cz\end{aligned}$$

Statistical models



Hypotheses

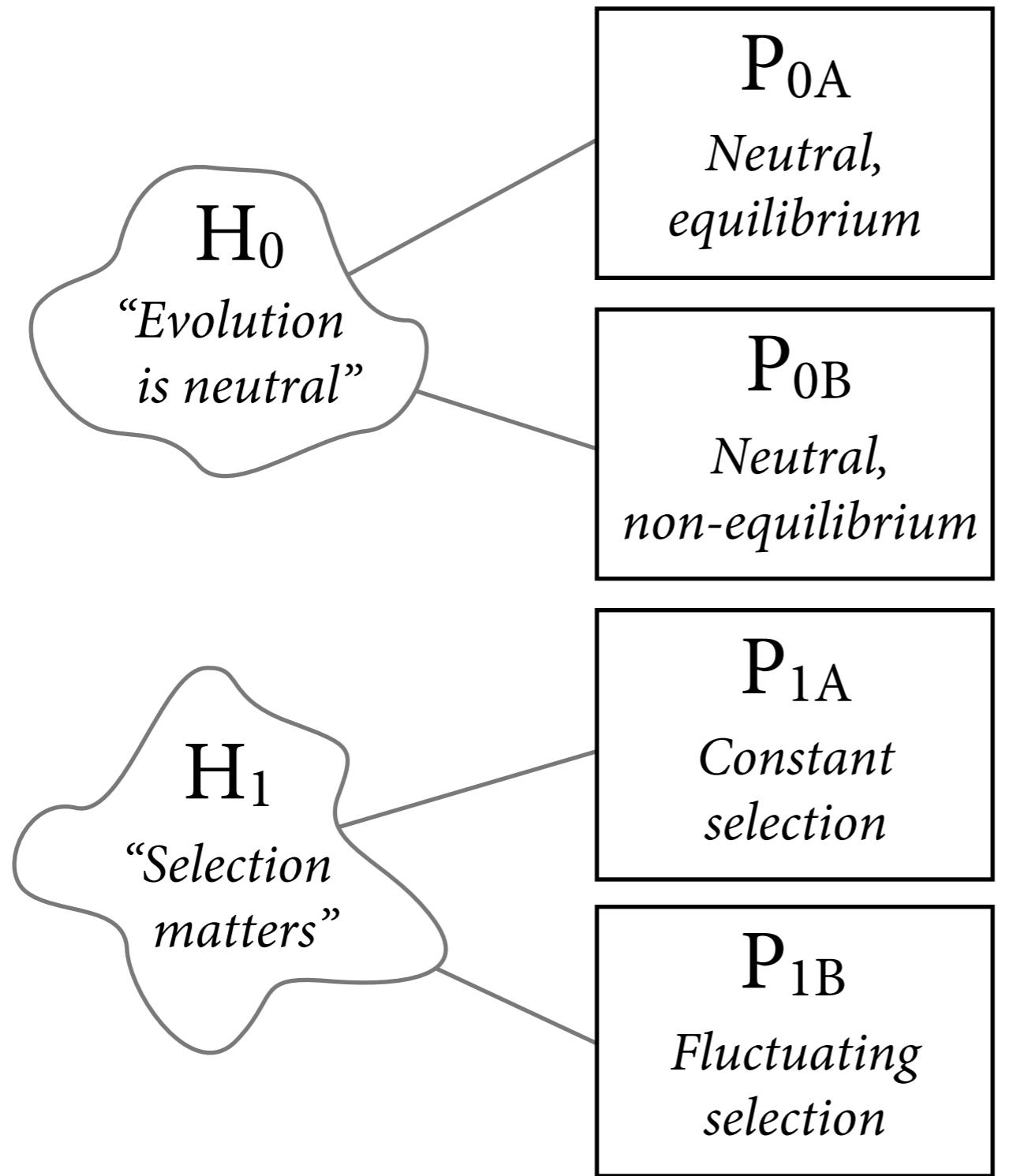
Process models

Statistical models

H_0
*“Evolution
is neutral”*

H_1
*“Selection
matters”*

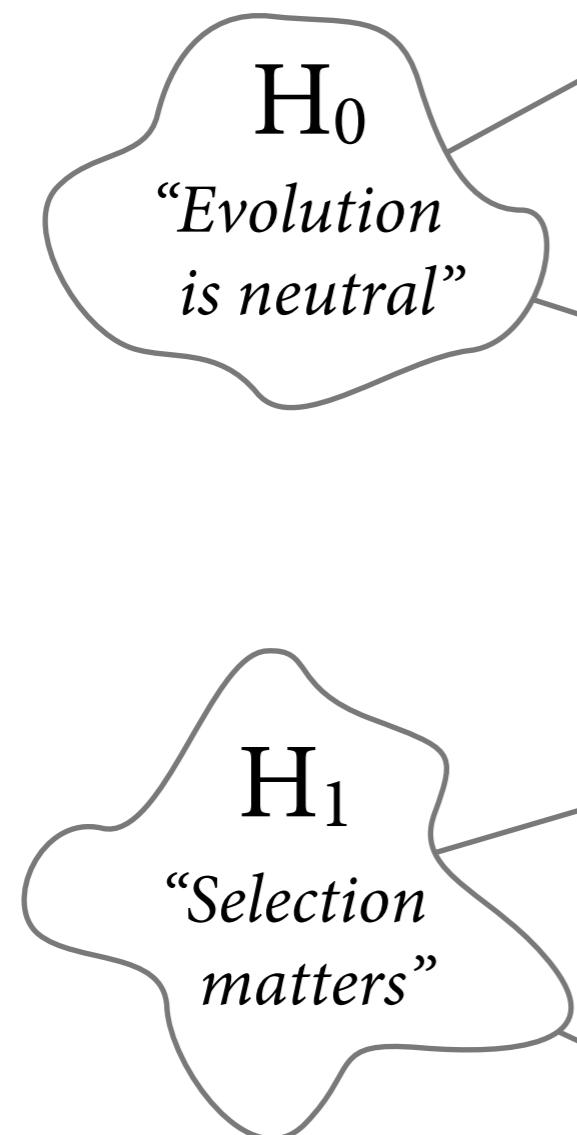
Hypotheses



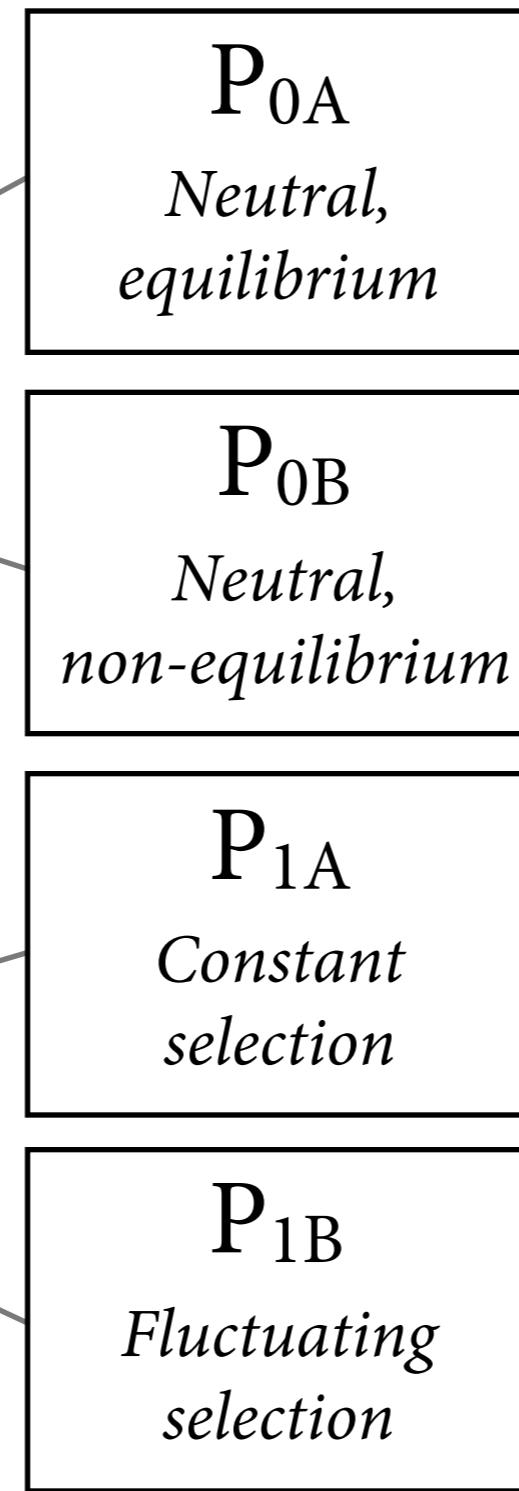
Process models

Statistical models

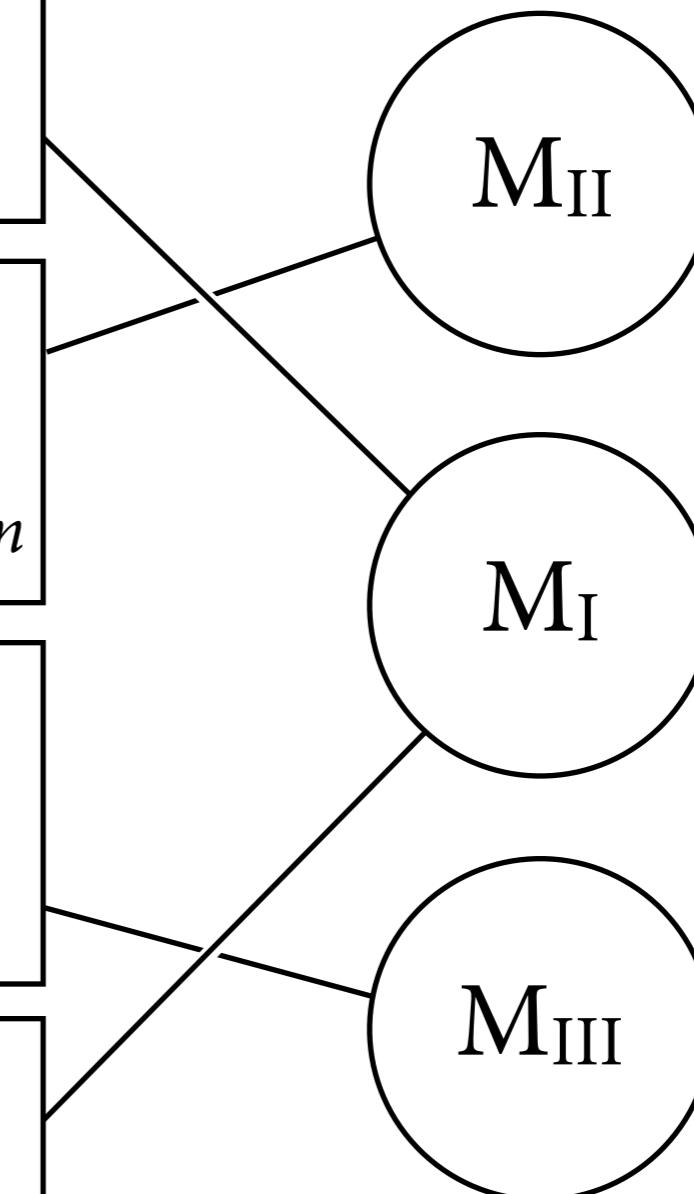
Hypotheses



Process models



Statistical models

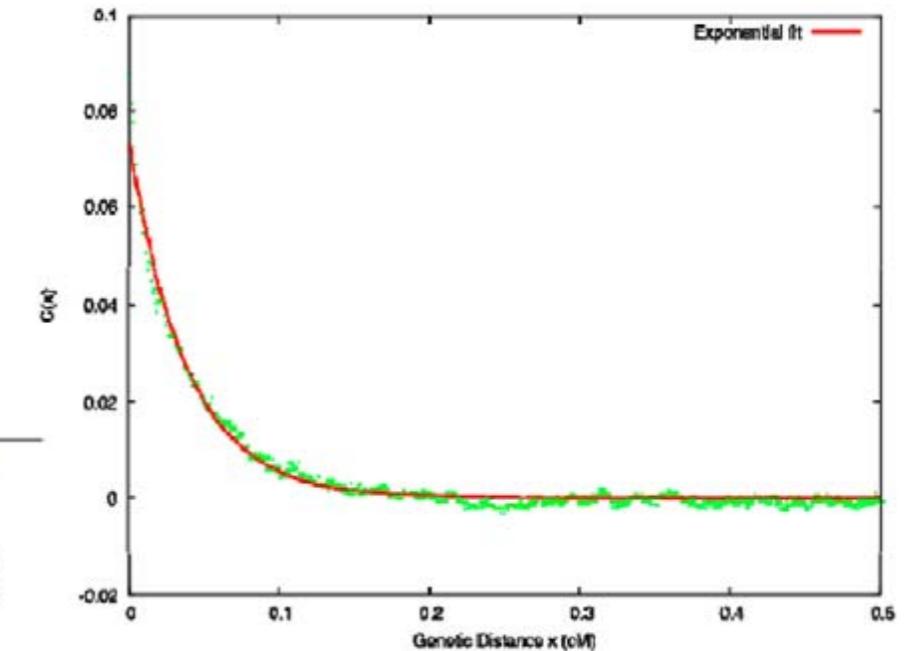
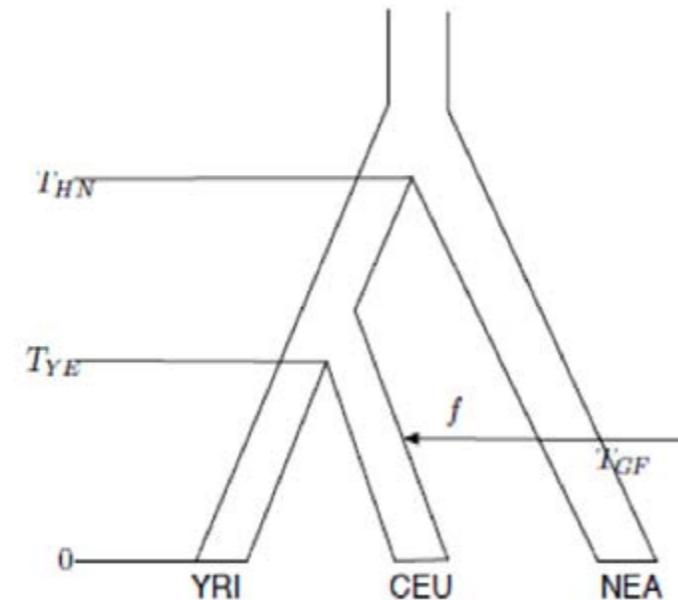




The Date of Interbreeding between Neandertals and Modern Humans 2015

Sriram Sankararaman^{1,2*}, Nick Patterson², Heng Li², Svante Pääbo^{3*}, David Reich^{1,2*}

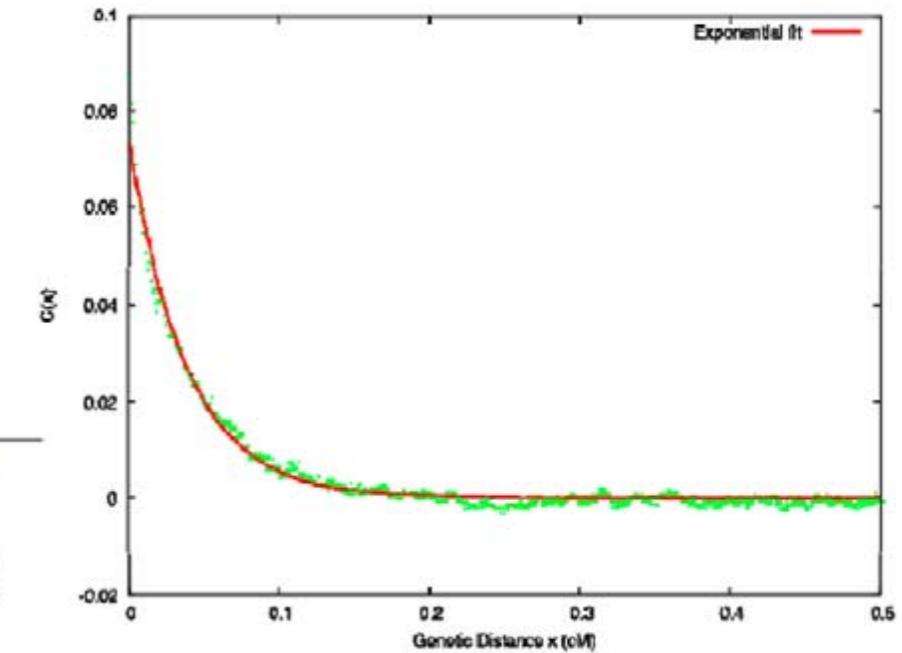
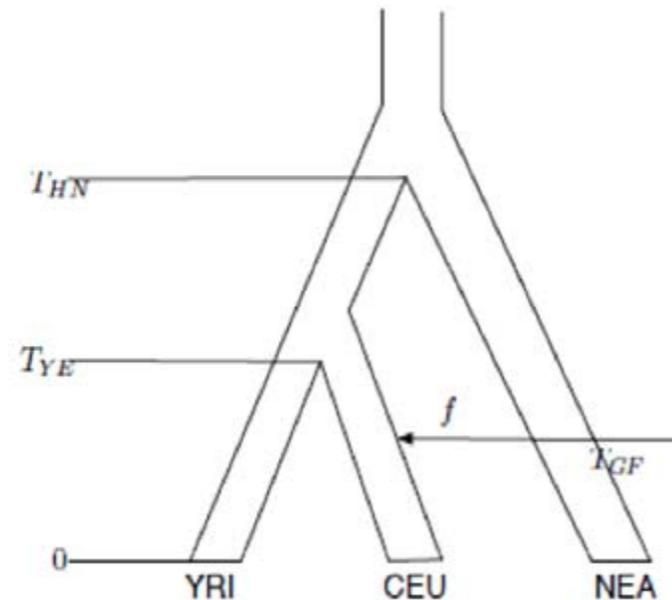
(a) Gene flow <100,000 years ago



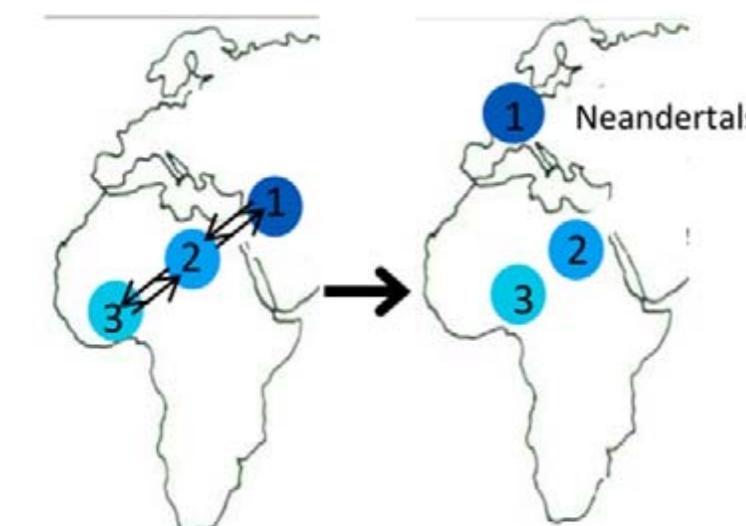
The Date of Interbreeding between Neandertals and Modern Humans 2015

Sriram Sankararaman^{1,2*}, Nick Patterson², Heng Li², Svante Pääbo^{3*}, David Reich^{1,2*}

(a) Gene flow <100,000 years ago

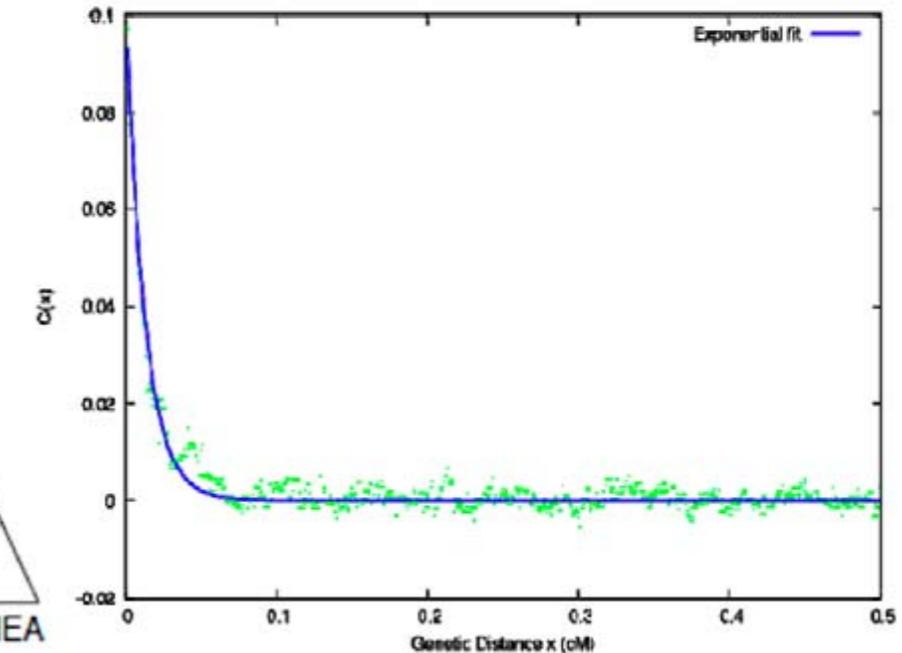
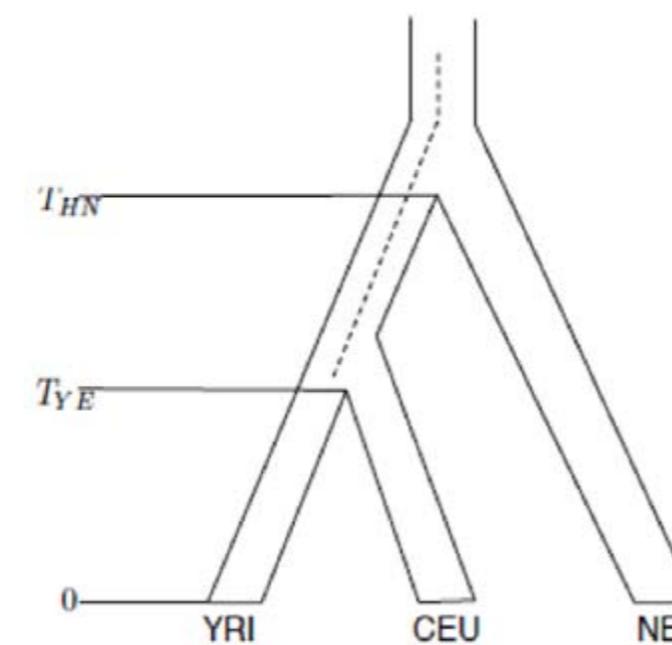


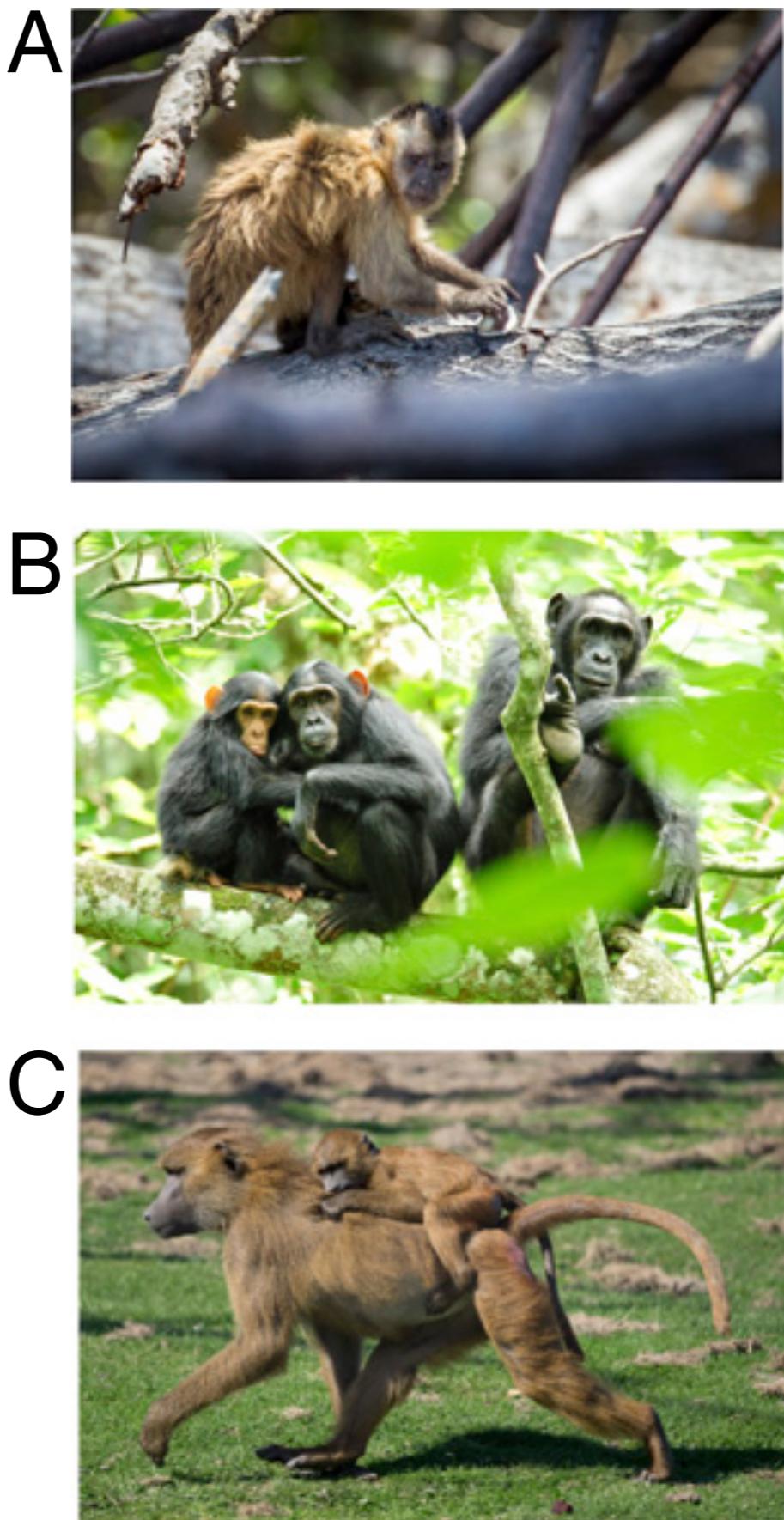
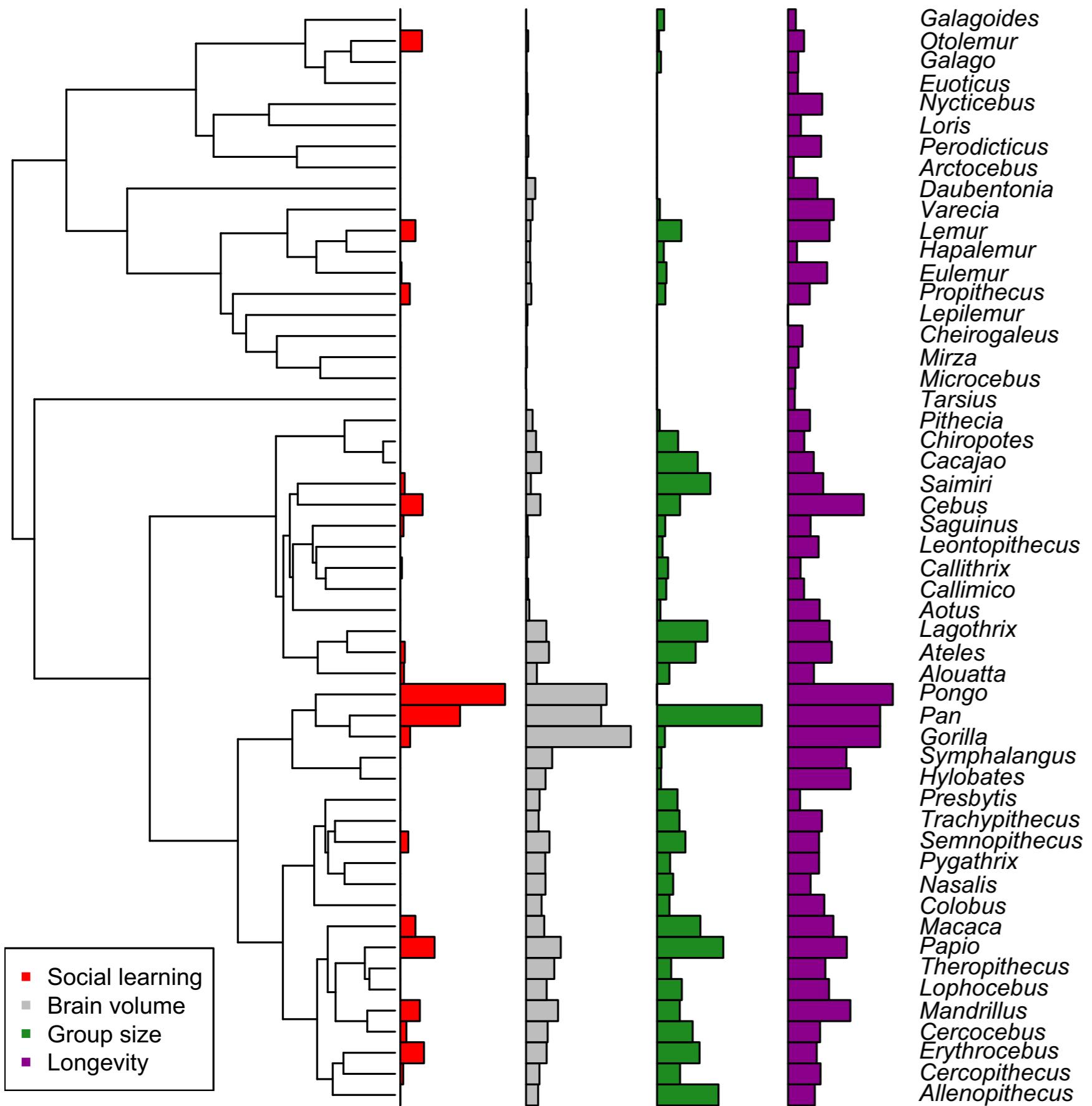
(b) Ancient structure >230,000 years ago



Human structure
to Neandertal-
modern human
divergence

Northeast
Africans retain
extra proximity to
Neandertals





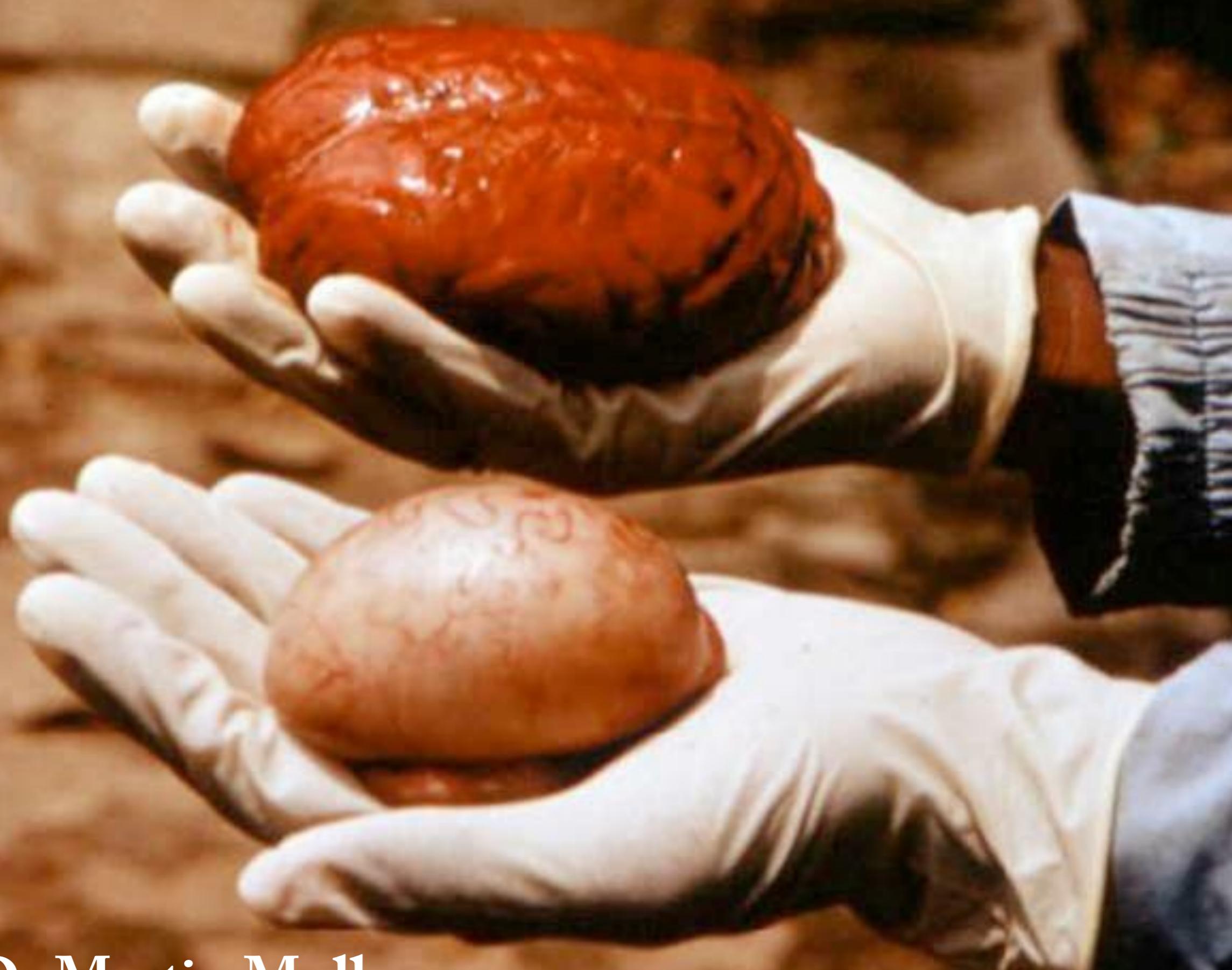
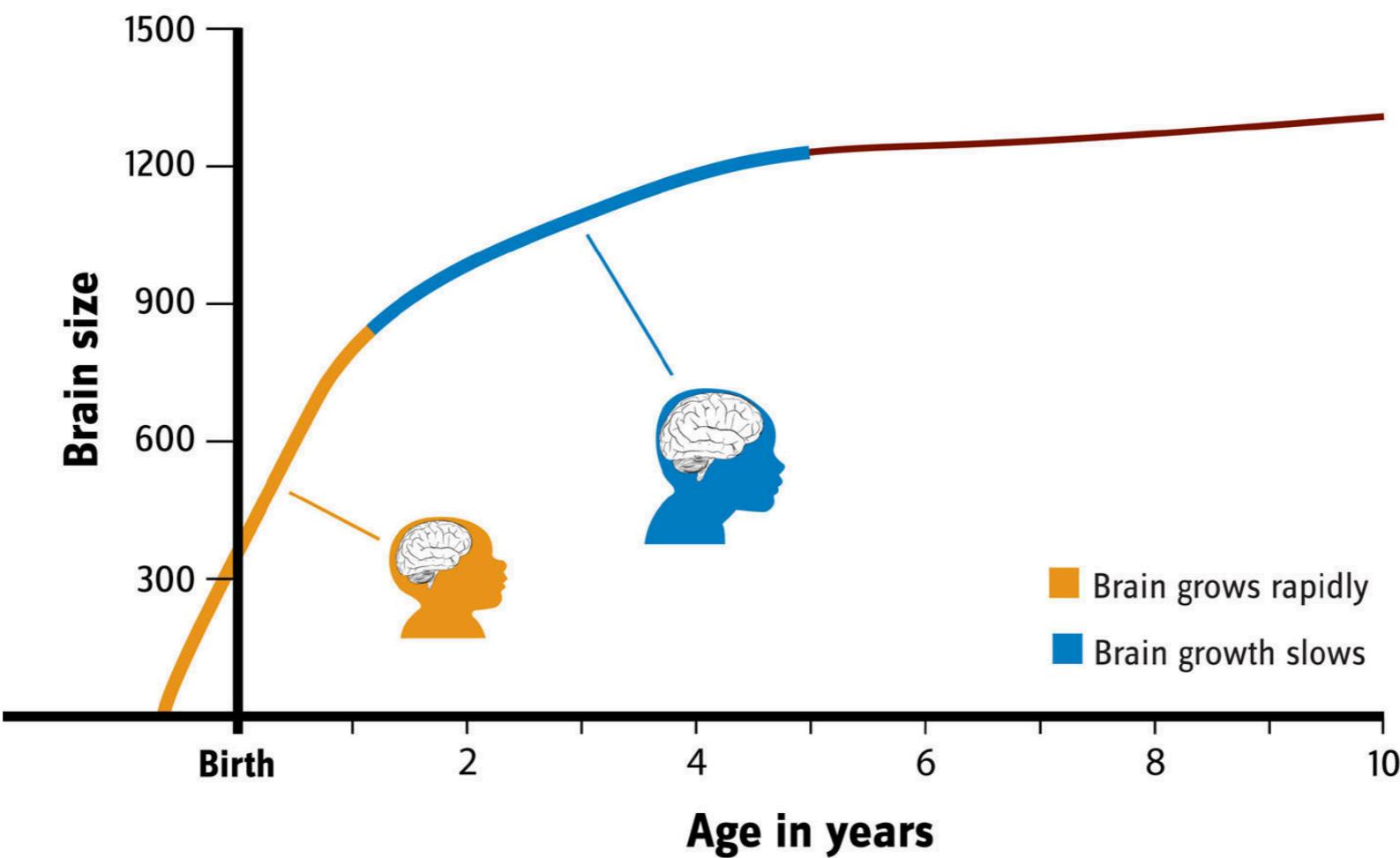
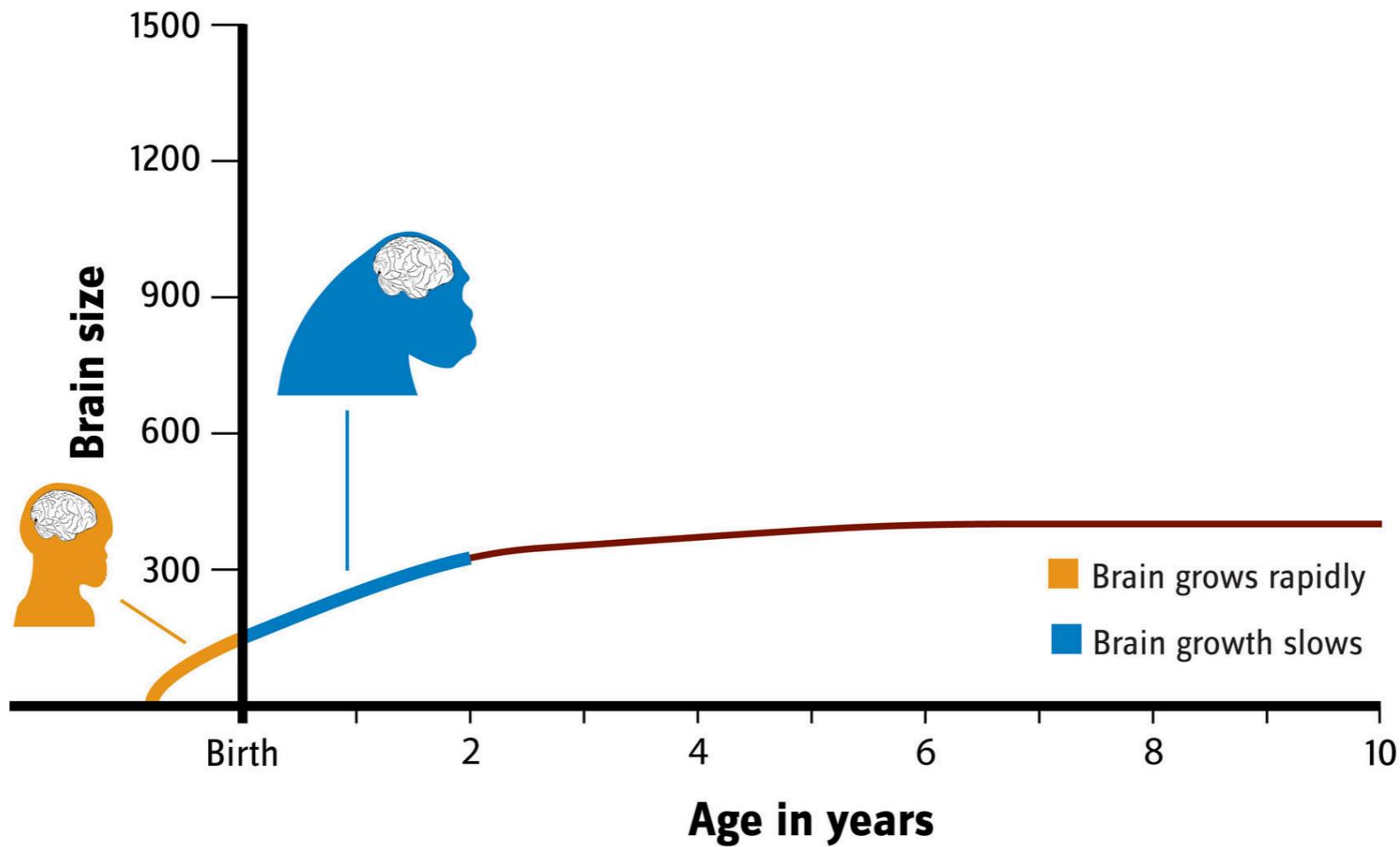
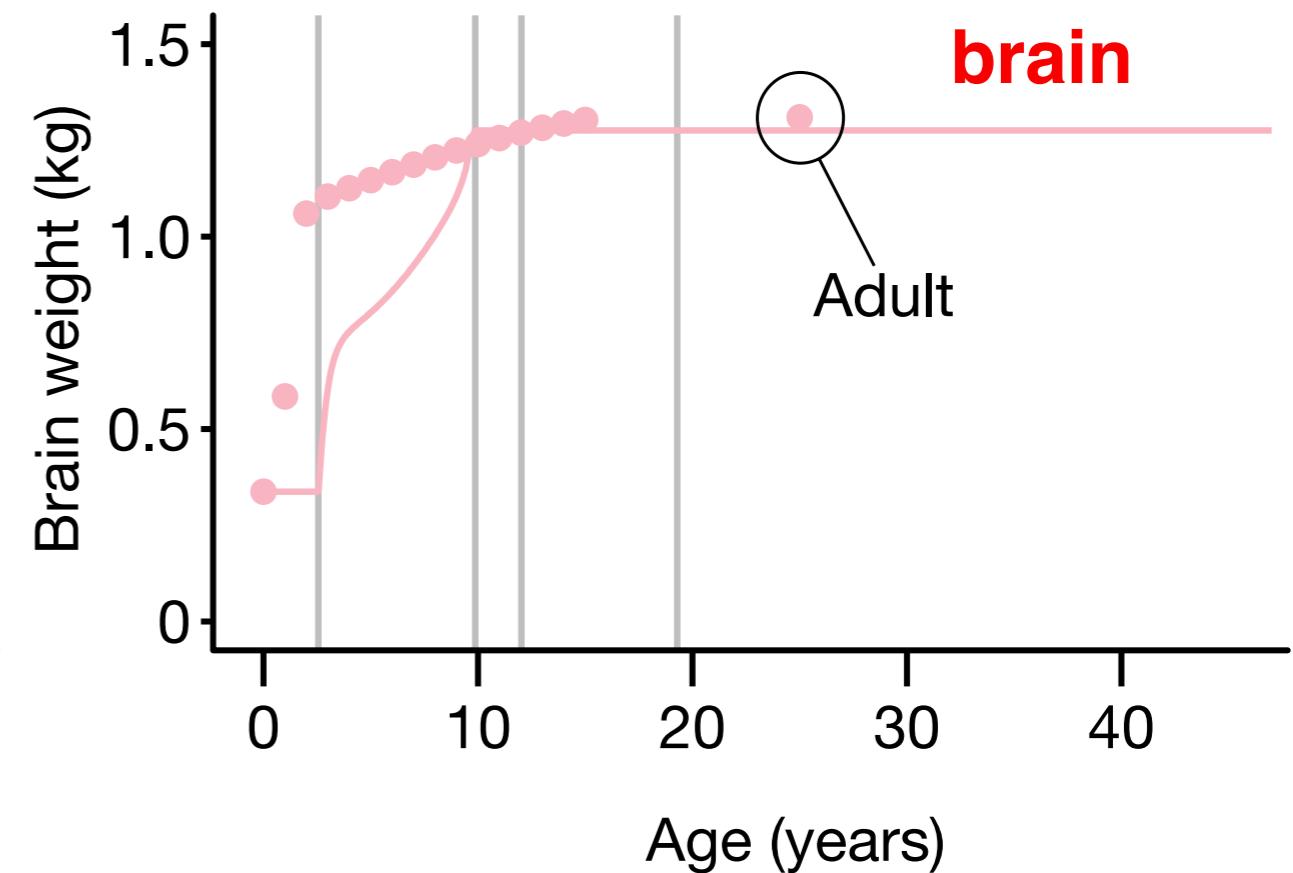
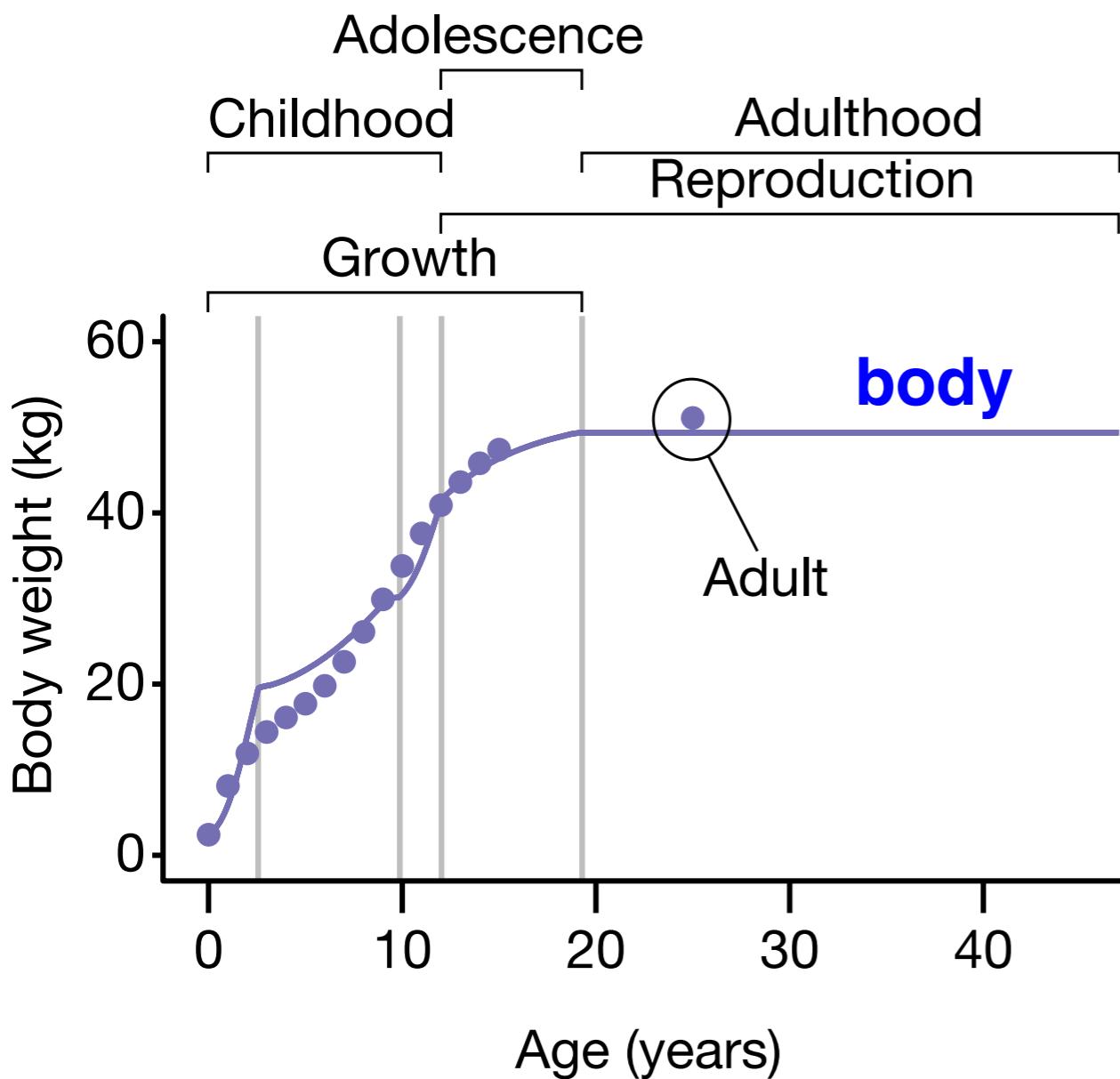


Photo: Dr Martin Muller





González-Forero & Gardner 2018.
Ecological and social drivers of human brain-size evolution



Brian Wood









JACK TAR ON A SPREE—"LAYING IN PROG." SKETCHED BY OUR SPECIAL ARTIST.

A delightful book on the recognition, gathering, preparation and use of the natural health foods that grow wild all about us. The lore contained here can turn every field, forest, swamp and vacant lot and roadside into a health-food market with free merchandise.

stalking the wild asparagus

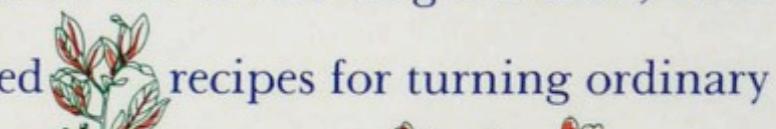
by Euell Gibbons

With a Remembrance of the Author by John McPhee

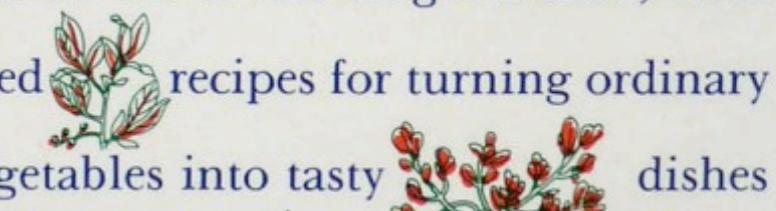
This is a book for lovers of good food, with hundreds of detailed recipes for turning ordinary wild fruits and vegetables into tasty dishes that would grace any table. The information contained in these pages will enable you to save many times the price on grocery bills and diet supplements. This book offers you economy, delicious food, fun and healthy activity.



recipes



wild fruits and vegetables into tasty



dishes



that would grace any table.



The information



contained in these pages will enable you to



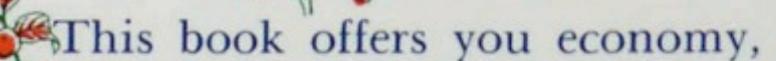
save many times the price on grocery bills and



diet supplements.



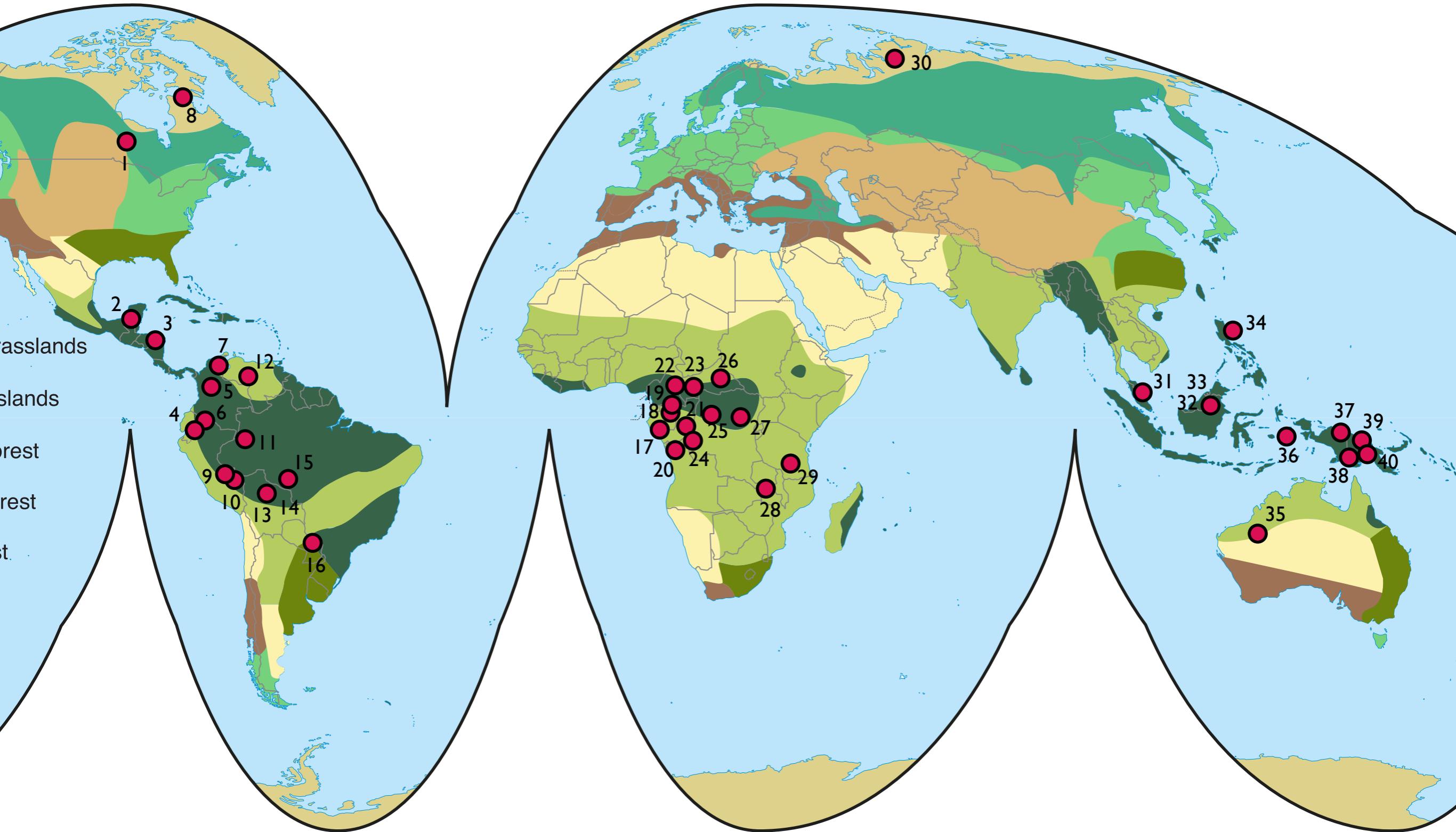
This book offers you economy,



delicious food, fun and healthy activity.



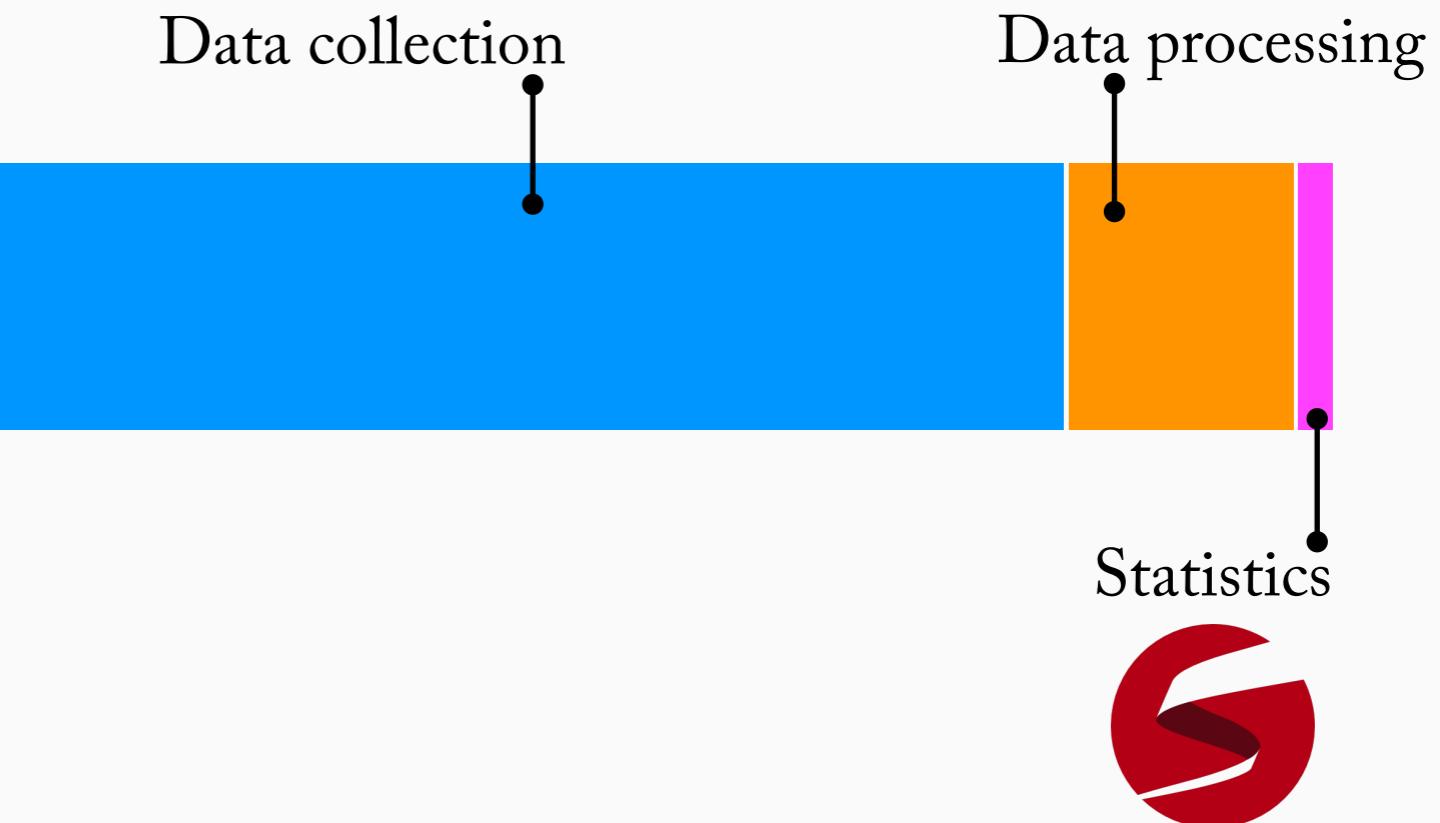
THE LIFE HISTORY OF HUMAN FORAGING: CROSS-CULTURAL AND INDIVIDUAL VARIATION

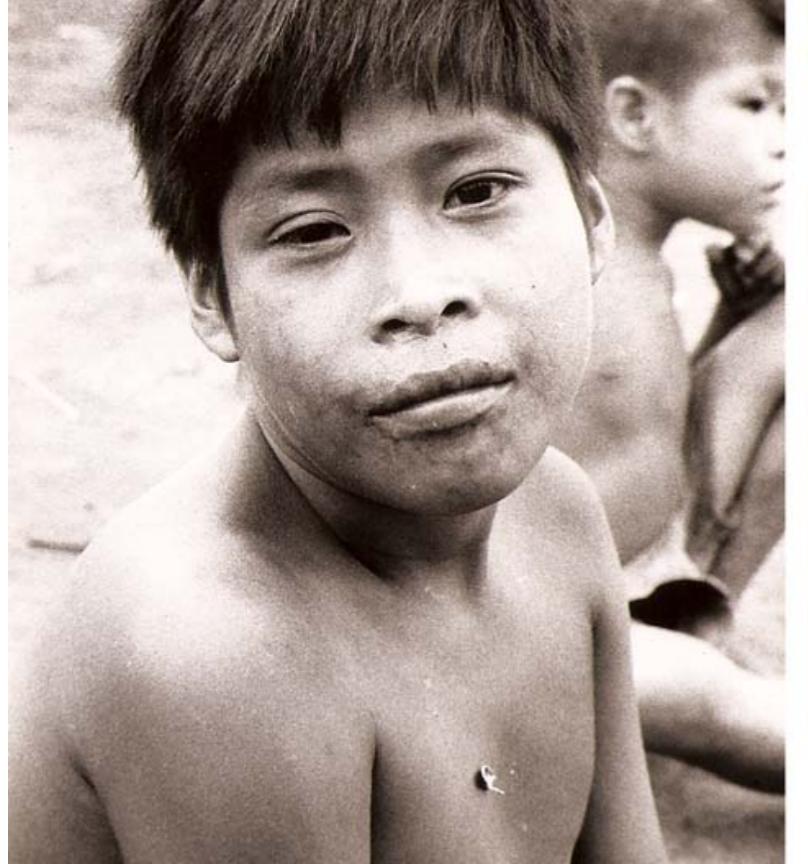


THE LIFE HISTORY OF HUMAN FORAGING: CROSS-CULTURAL AND INDIVIDUAL VARIATION

JEREMY KOSTER^{1,2}, RICHARD MCELREATH², KIM HILL, DOUGLAS W. YU, GLENN H. SHEPARD JR., NATHALIE VAN VLIET ET AL, MICHAEL GURVEN, HILLARD KAPLAN, BENJAMIN TRUMBLE, REBECCA BLIEGE BIRD, DOUGLAS BIRD, BRIAN CODDING, LAUREN COAD, LUIS PACHECO, BRUCE WINTERHALDER, KAREN D. LUPO, DAVE N. SCHMITT, PAUL SILLITOE, MARGARET FRANZEN, MICHAEL ALVARD, VIVEK VENKATARAMAN, THOMAS KRAFT, KIRK ENDICOTT, STEPHEN BECKERMAN, STUART A. MARKS, THOMAS HEADLAND, MARGARETHA PANGAU-ADAM ET AL, ANDERS SIREN, KAREN KRAMER, RUSSELL GREAVES, VICTORIA REYES-GARCÍA, MAXIMILIEN GUÈZE, ROMAIN DUDA, ÁLVARO FERNÁNDEZ-LLAMAZARES, SANDRINE GALLOIS, LUCENTEZZA NAPITUPULU, ROY ELLEN, JOHN ZIKER, MARTIN R. NIELSEN, ELSPETH READY, CHRISTOPHER HEALEY, AND CODY ROSS

Dr Jeremy Koster (left)

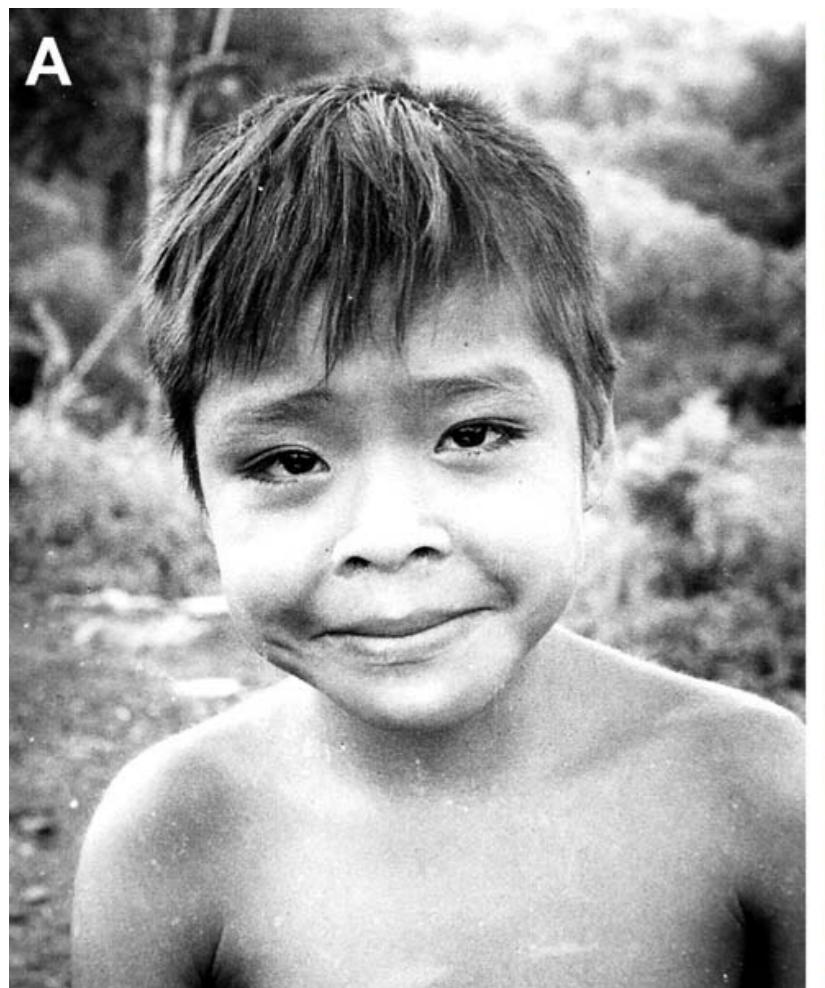




1978



2009



From:

Can Anthropologists Distinguish Good and Poor Hunters? Implications for Hunting Hypotheses, Sharing Conventions, and Cultural Transmission

Kim Hill and Keith Kintigh

Life History of Production Skill

- Goals
 - Estimate skill development
 - Take variation seriously
 - Develop stats machinery
- Sample
 - 40 research sites
 - 1821 individuals
 - 21,160 foraging trips
 - 23,747 harvests
 - Uncountable headaches



Bad Data Headaches

- Missing values: age, duration, technology
- Uncertainty: age bins, age estimates
- Massive imbalance in sampling



“Pooh?” said Piglet.

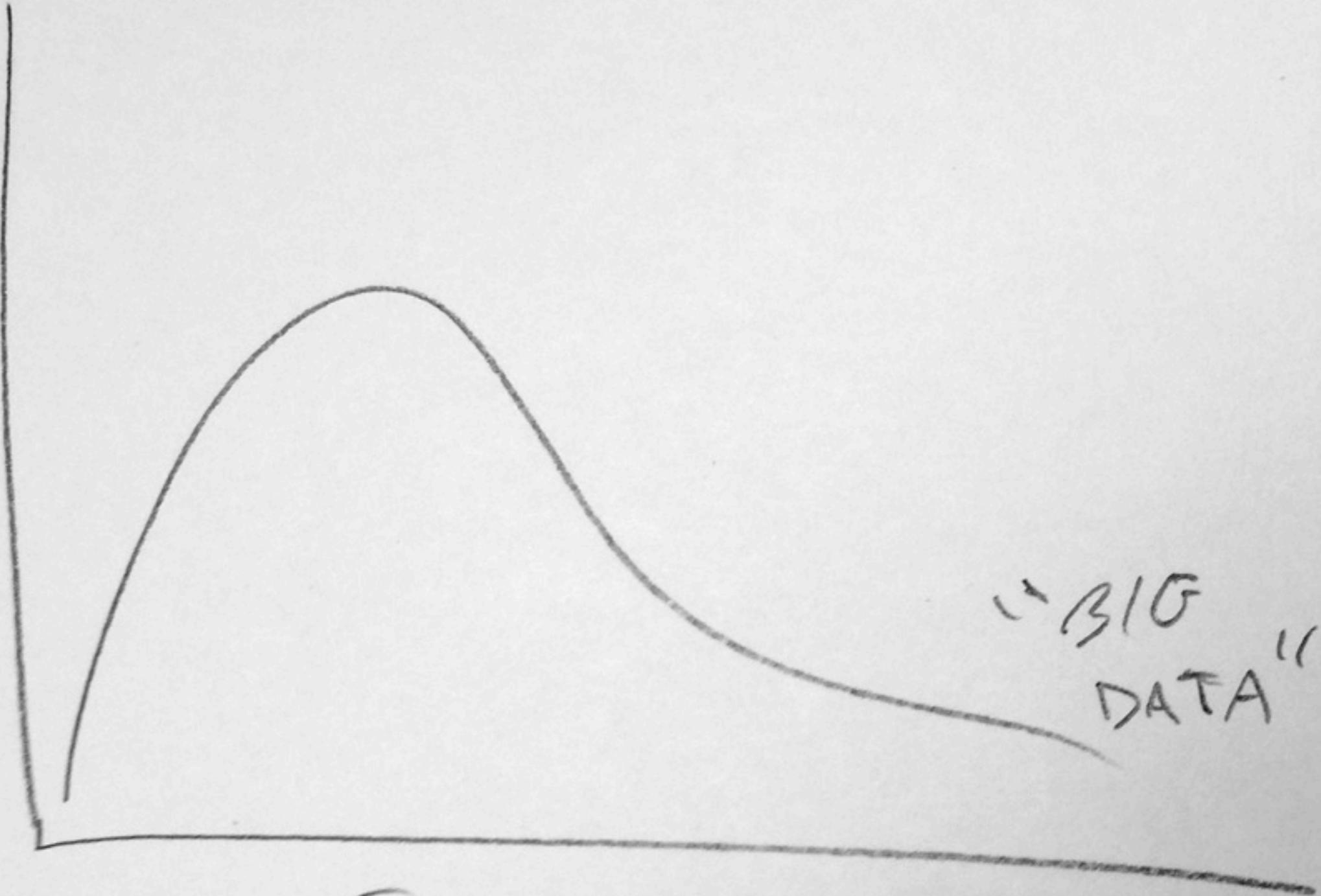
“Yes, Piglet?” said Pooh.

“27417 parameters,” said Piglet.

“Oh, bother,” said Pooh.

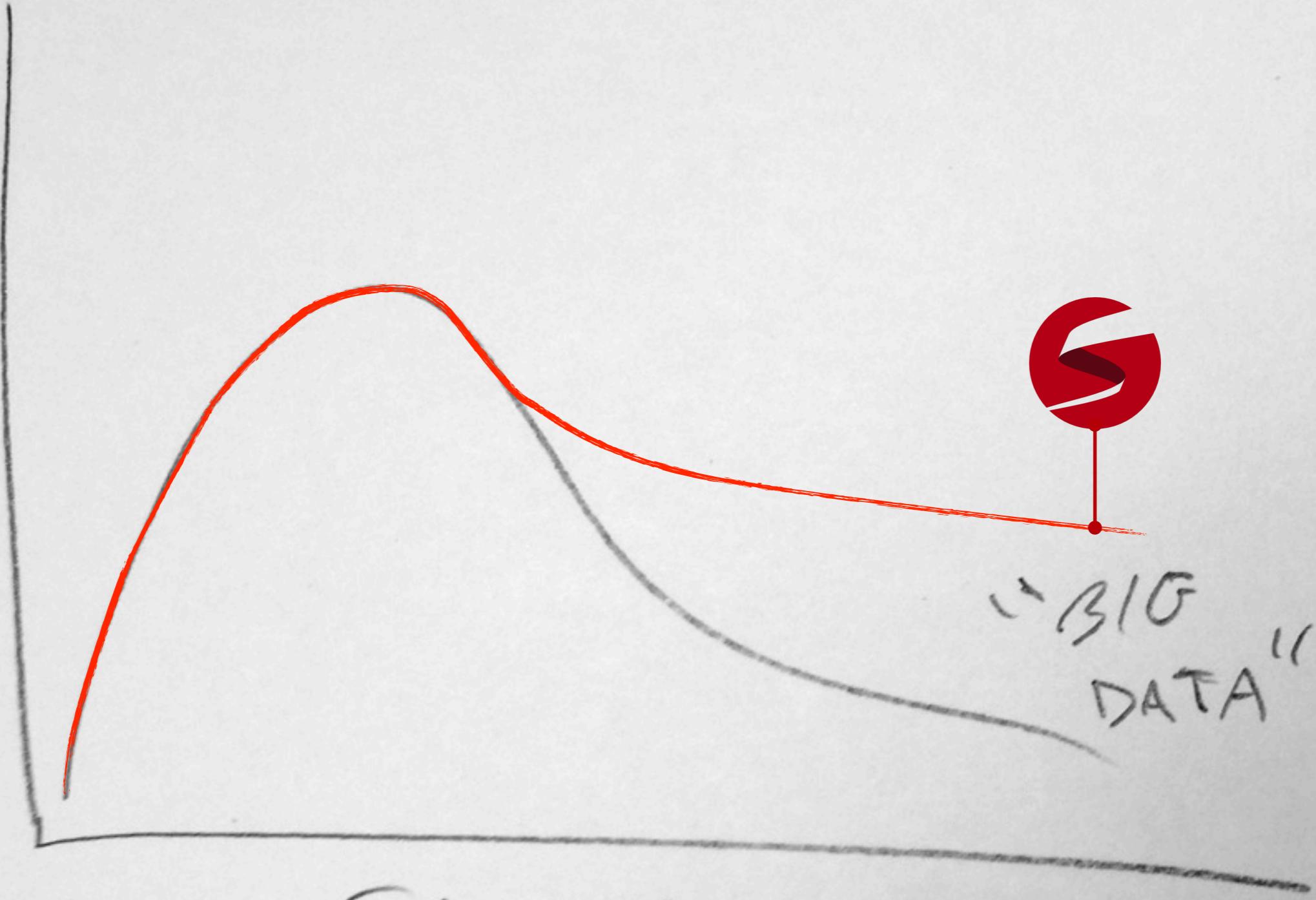


Model
Complexity



SAMPLE SIZE

Model
Complexity



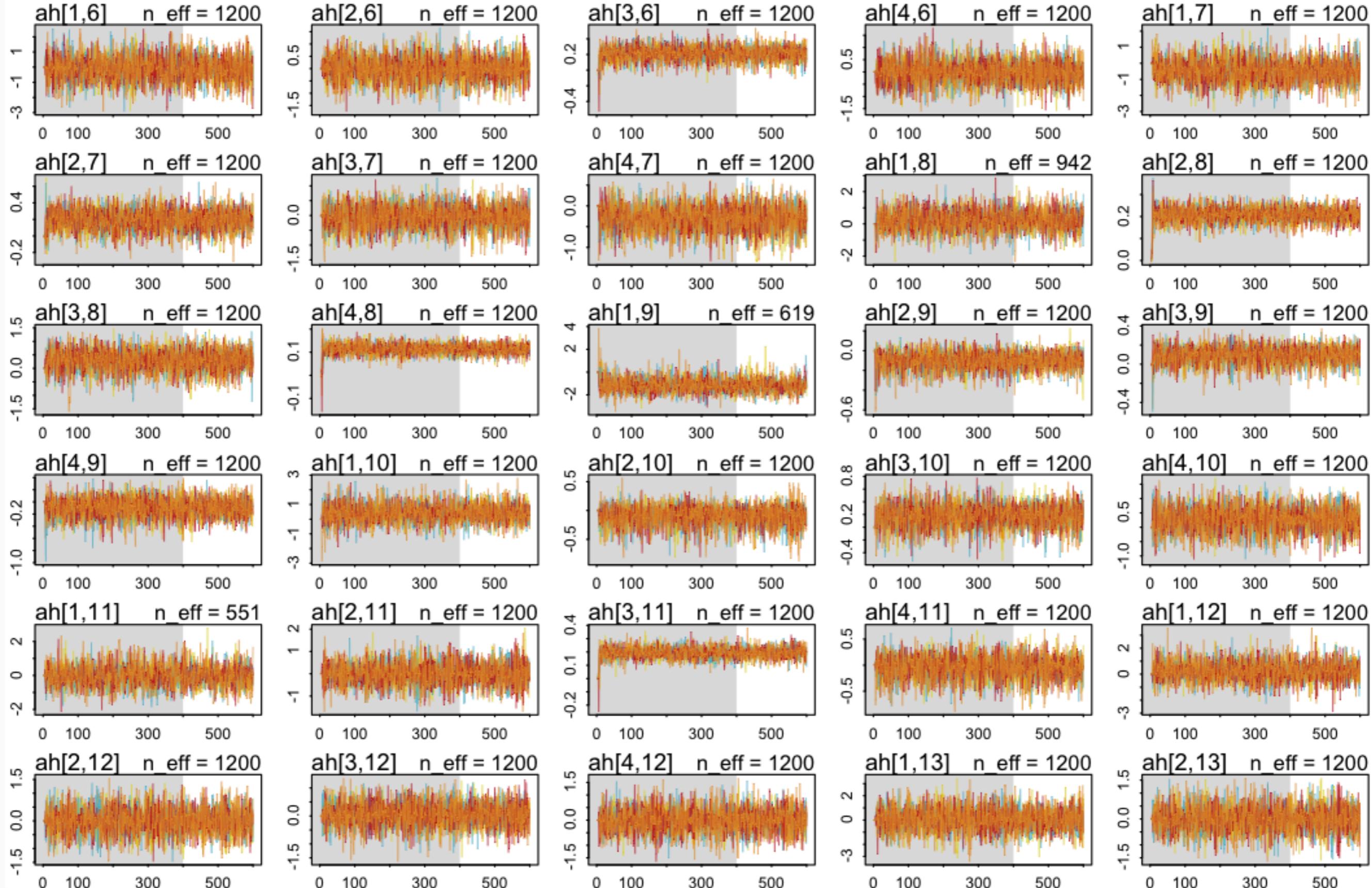


Scaffolds

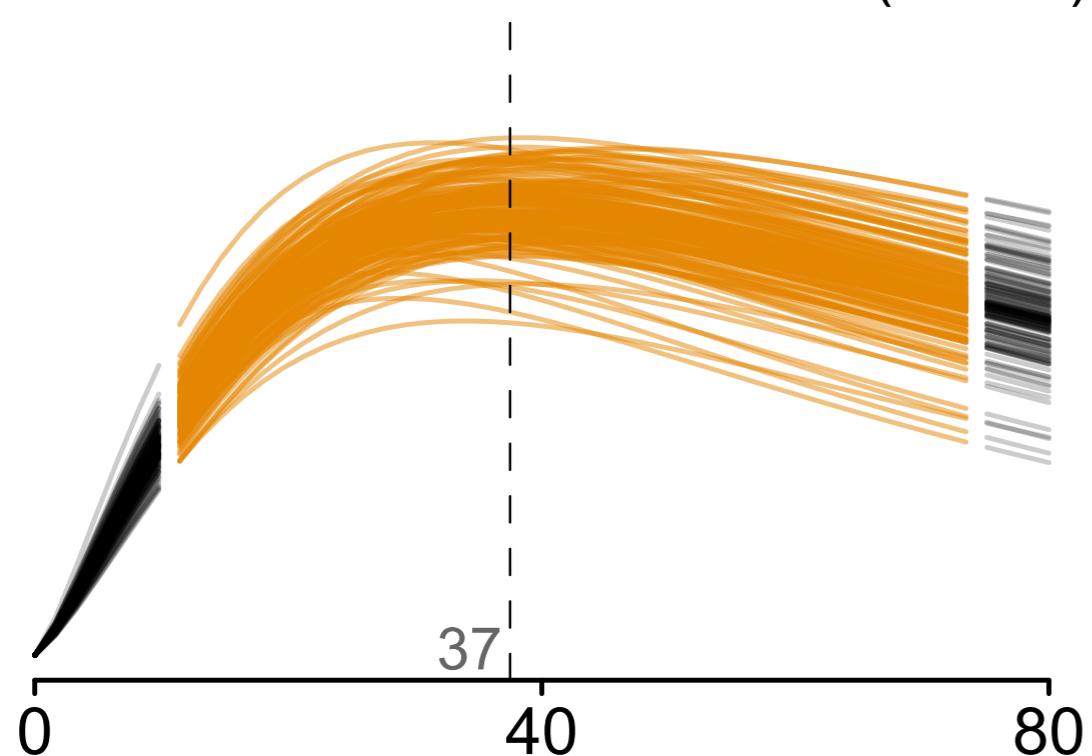
- A month of modeling & fake data simulation (priors)
- Submit grant (wait for it...), get grant!
- Develop first Stan model (multilevel)
- Simulate missingness and uncertainty
- Develop second Stan model (imputation, marginalization)
- Analyze real sample

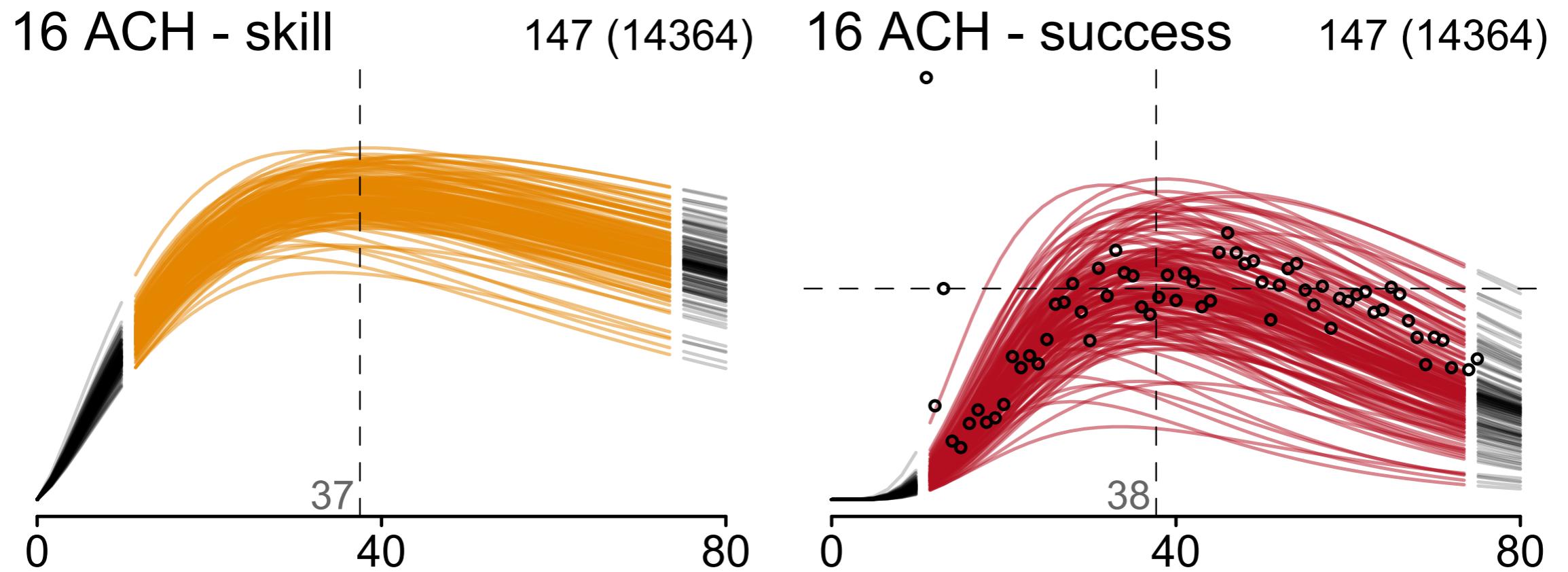


What Happened

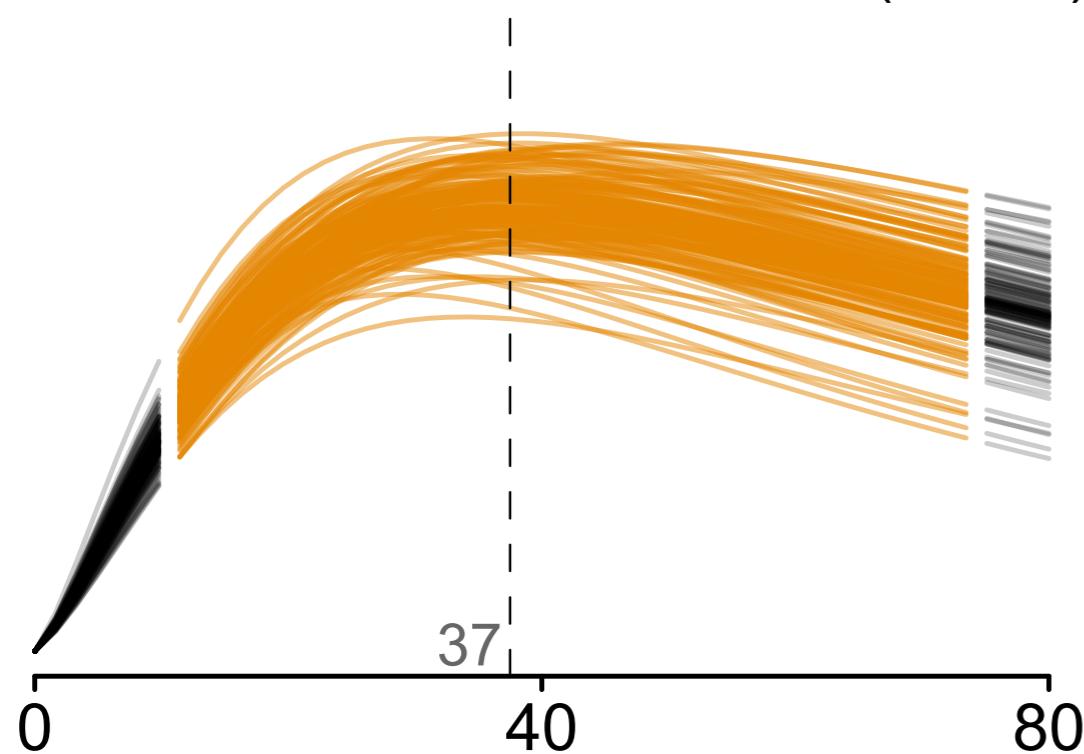


16 ACH - skill



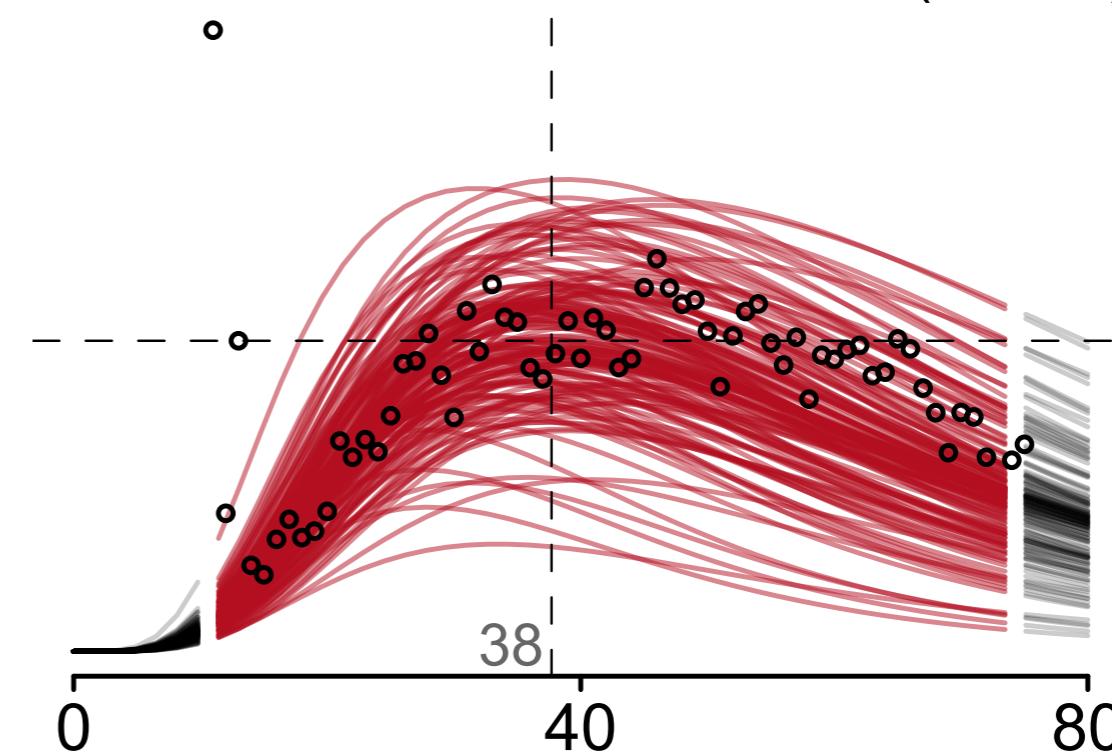


16 ACH - skill



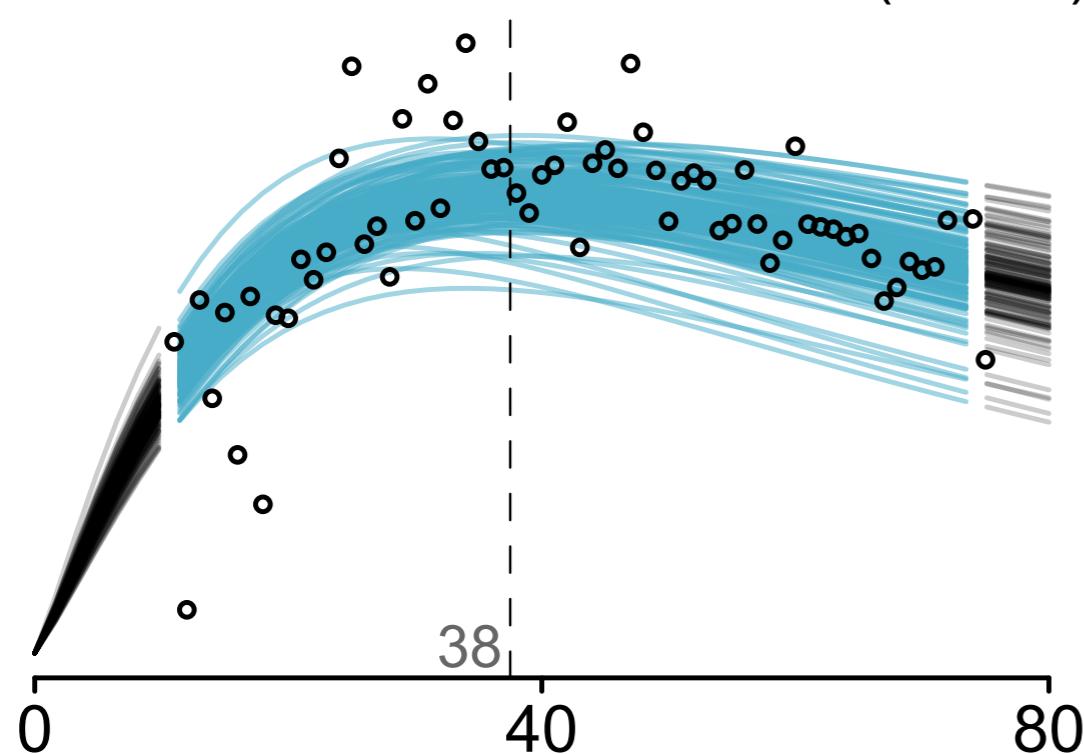
147 (14364)

16 ACH - success

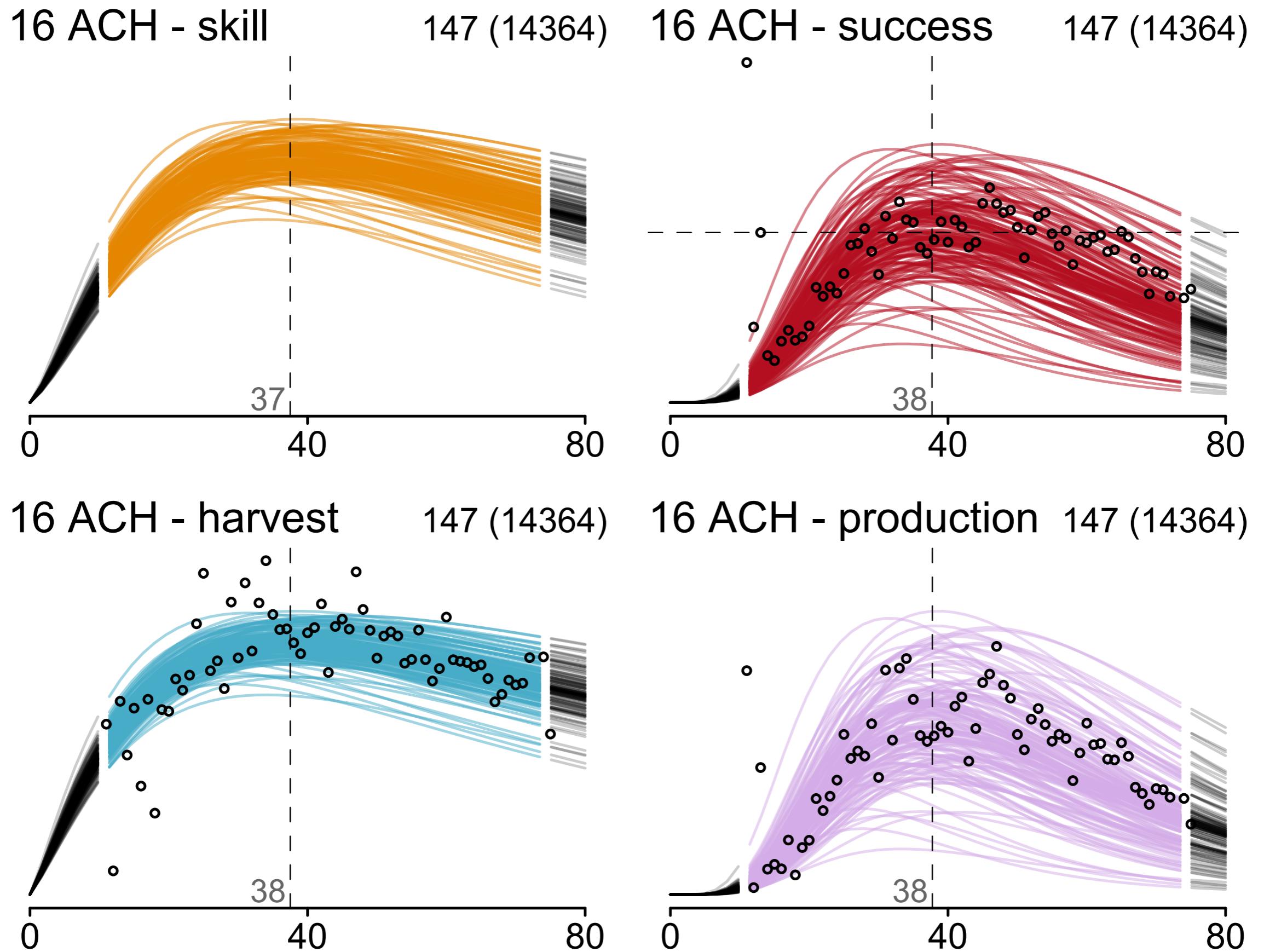


147 (14364)

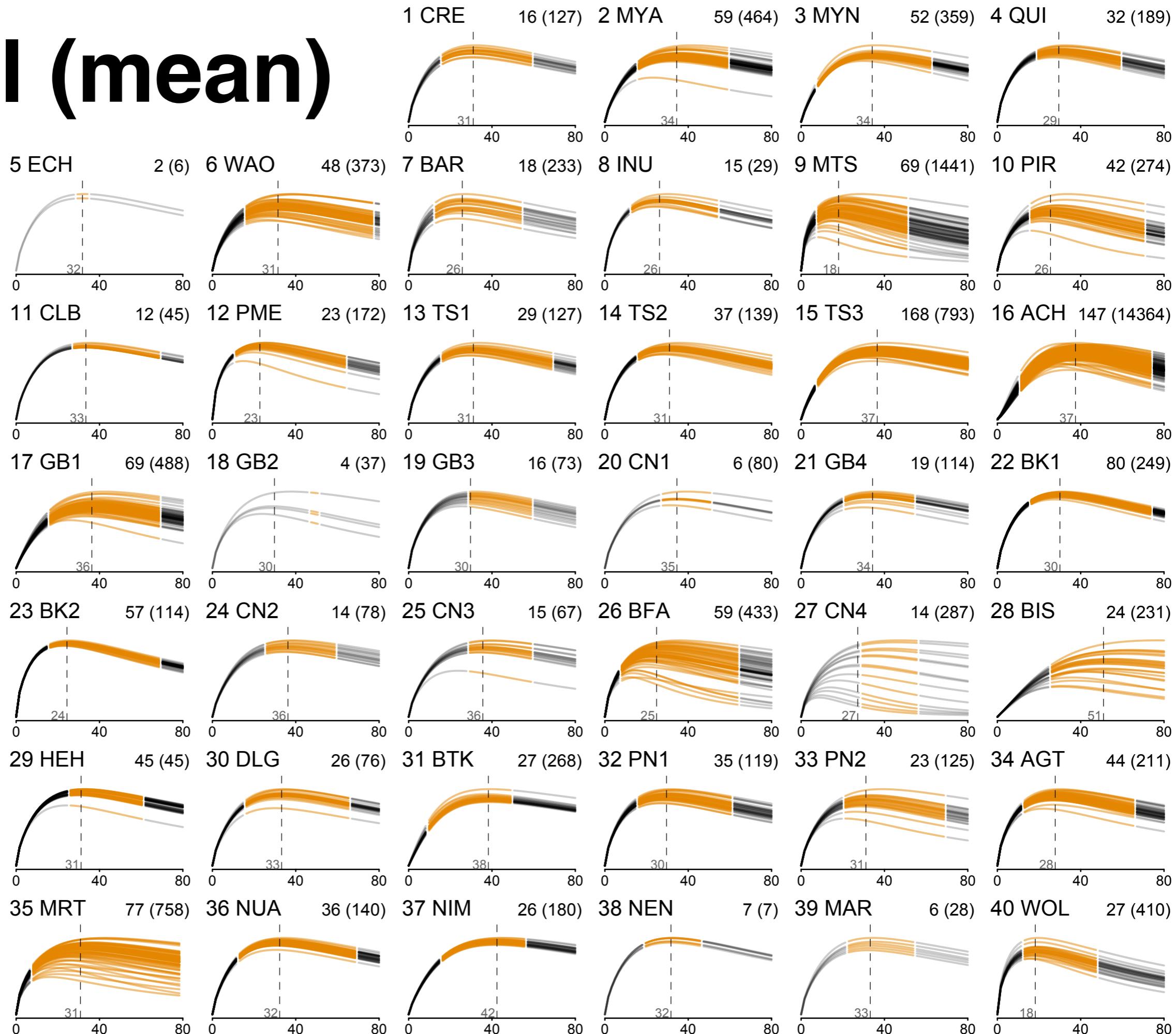
16 ACH - harvest



147 (14364)



Skill (mean)



Bad Data in Search of Good Home

- Evolution is not your friend
- Not experimental science
- Null models useless
- Data scarce, not simple
- Flexible, scalable, open tools
- Community crucial



