

# 北京师范大学心理学部

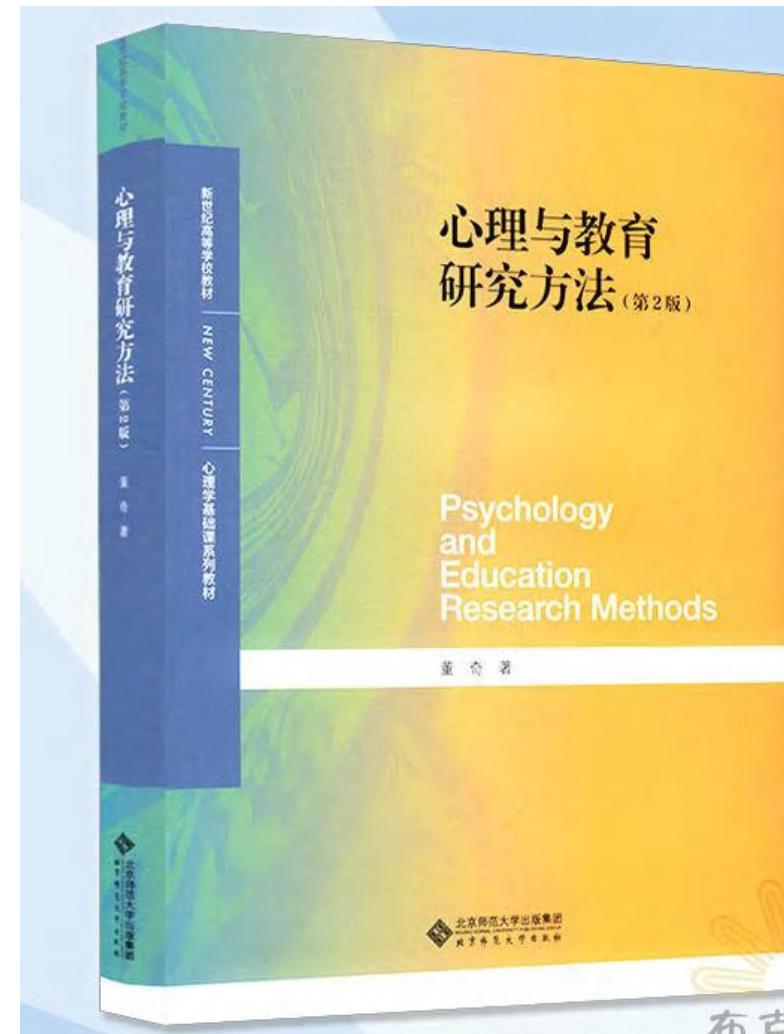
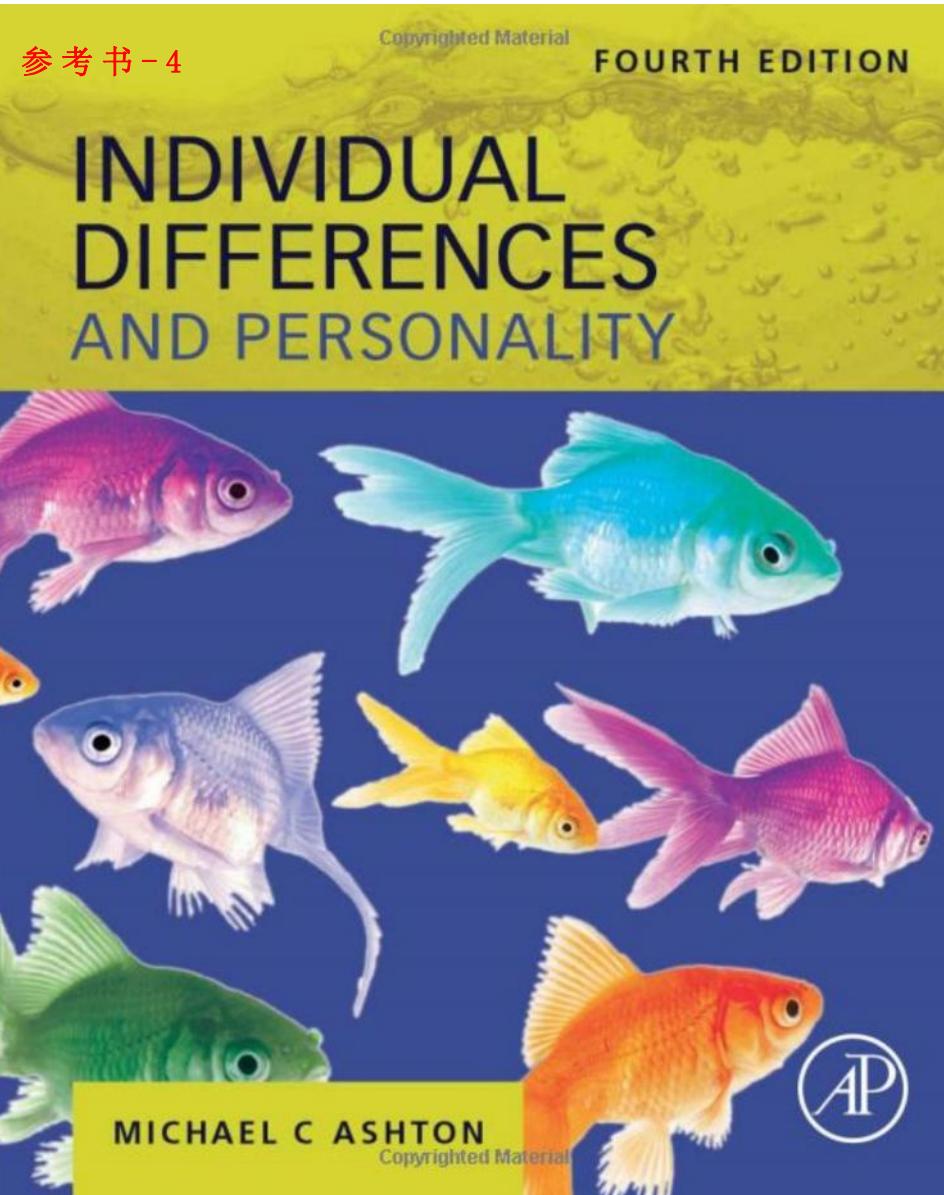
## Developmental Population Neuroscience

发展人口神经科学（个体差异的测量理论）

左西年 (Xi-Nian Zuo)

Beijing Normal University  
State Key Lab of Cognitive Neuroscience & Learning

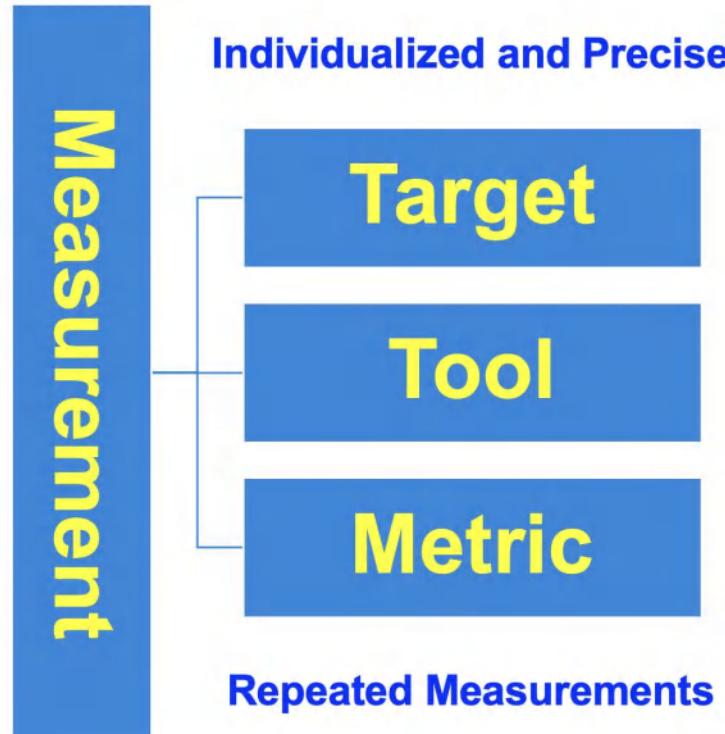
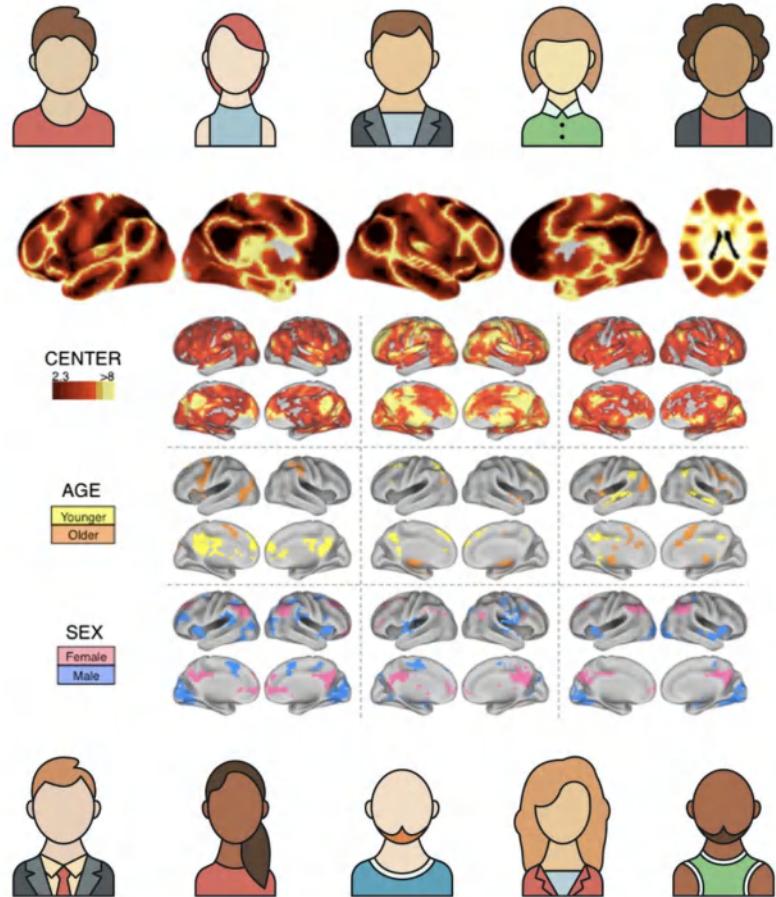
National Basic Science Data Center  
Chinese Data-sharing Warehouse for In-vivo Imaging Brain



心理研究与方法  
第二版

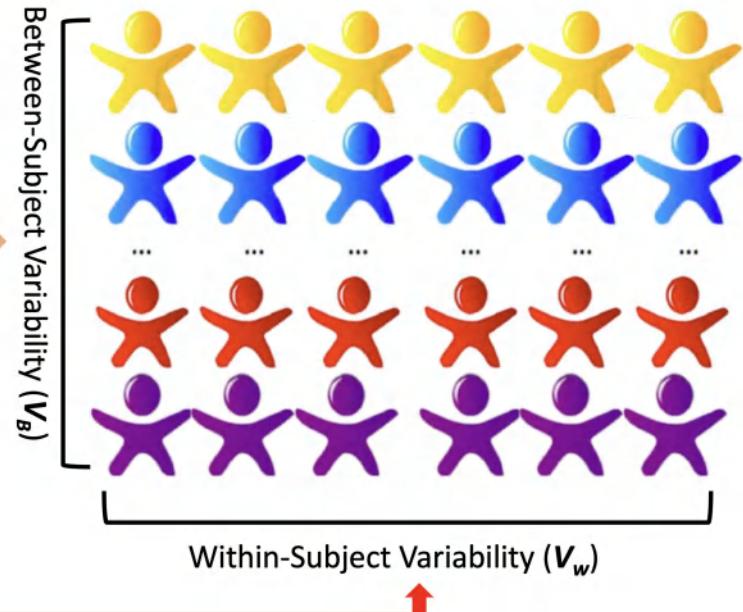
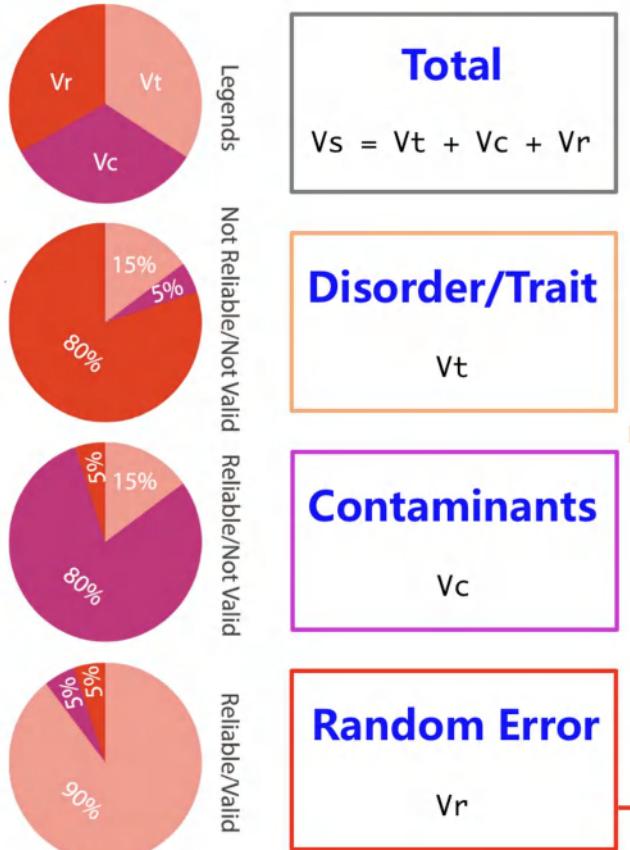
心理学专业经典教材

# Individual Differences: Measurement Theory



个体差异的精准测量是科学实现个体化与精准化转化的前提。理论上，个体差异研究需要重复测量，测量包括三个方面内涵：**测量目标、测量工具和测量指标**，研发评估一项测量需要从上述三个方面综合考虑。

# Measurement Theory: Reliability versus Validity



*Nature Human Behaviour* (2019); *Frontiers in Neuroscience* (2019)

“工欲效其事，必先信其器”——个体差异的科学靠谱吗？

邢秀侠<sup>1</sup>, 左西年<sup>2\*</sup>

<sup>1</sup> 北京工业大学应用数理学院, 北京 100124;

<sup>2</sup> 中国科学院心理研究所脑成像研究中心, 北京 100101

\*联系人, E-mail: zuoxn@psych.ac.cn

**“To do a valid job, must make tools reliable first”  
—A decent science of individual differences?**

Xiuxia Xing<sup>1</sup> & Xinian Zuo<sup>2\*</sup>

<sup>1</sup> College of Applied Sciences, Beijing University of Technology, Beijing 100124, China;

<sup>2</sup> Magnetic Resonance Imaging Research Center, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

doi: 10.1360/TB-2019-0270

作为人类特有的行为, 科学研究是社会文明的重要推动力量之一。近年来, 研究的可重复性问题成为科学关注的焦点。从心理科学到临床医学等领域, 研究的可重复性成为巨大挑战。生命科学研究的共同特点之一是对于测量工具的需求, 先进的技术会促进更为精准的测量。提升研究可靠性, 测量理论中的可信(可信度)或效度(效度)概念在不同学科都有涉及, 特别是在心理科学和医学中有明确的统计学界定, 但在其他学科未被充分认识, 尤其是交叉学科。

人与人之间为何表现出如此巨大的差别? 这是个体差异的科学问题。一直以来是生命科学领域备受关注的焦点。

*Nat Hum Behav* 在 2019 年 6 月 28 日在线发表题为

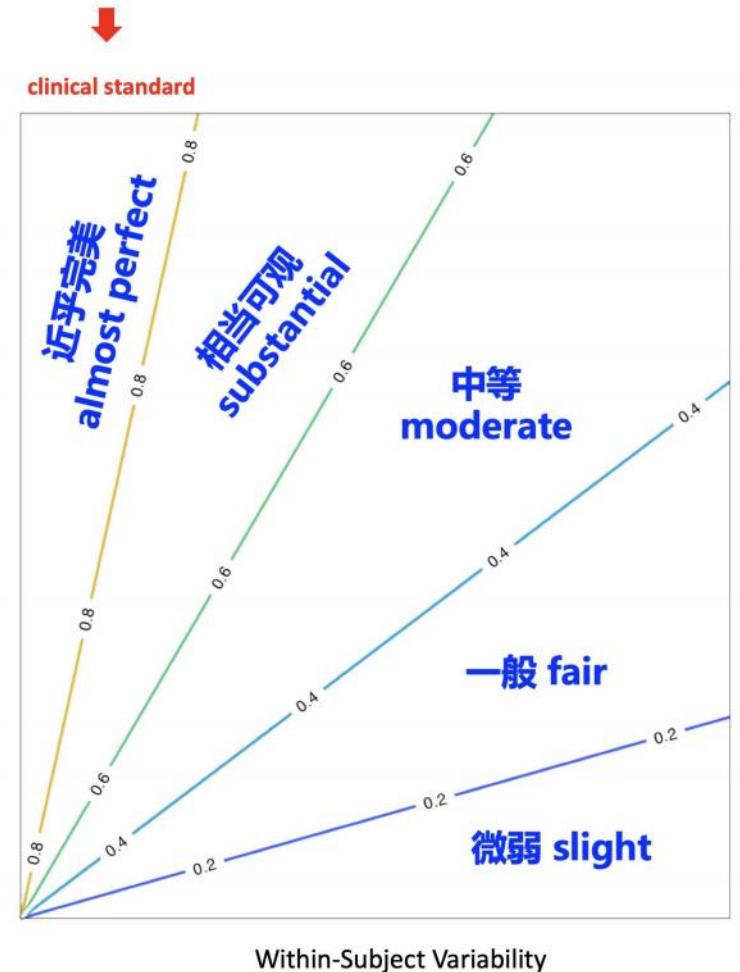
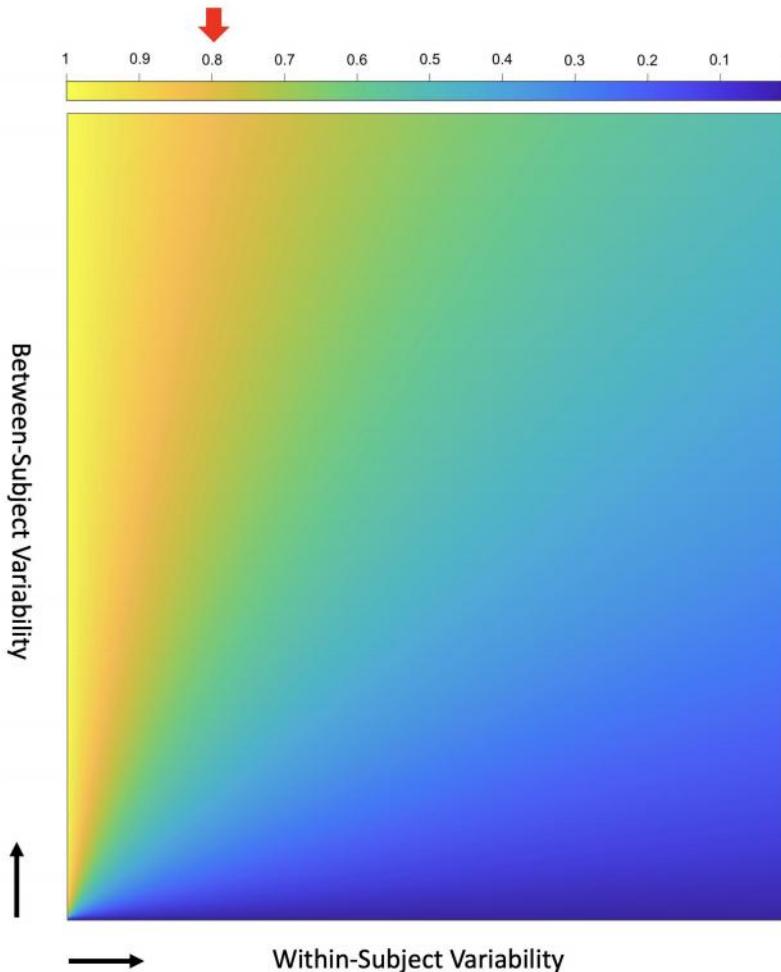
“Harnessing reliability for neuroscience research” 的评论文章<sup>[1]</sup>。以神经科学为例, 聚焦神经影像技术, 提出了个体差异测量效度统计学框架。在此框架下, 个体差异的总测量由三部分组成(图 1(a)): 研究对象特异的变化(疾病或特质测量), 研究对象非特异的变化(干扰和污染测量), 随机误差(随机噪声测量)。个体差异的测量采用重复测量设计(图 1(b))。图中相同颜色表示同一个人, 不同颜色表示不同的人。个体间差异测量包含“疾病或特质测量”与“干扰与污染测量”的总和, 而随机噪声测量是个体间差异测量。个体差异测量的信度是个体间差异测量在总测量中所占比例, 而疾病或特质特异变化所占比例则是个体差异测量的效度。由此, 测量的信度就像一个瓶子的盖子一样, 牢牢地限制住了测量的精度, 不可信的测量永远不可能有效; 与此同时, 测量的个体间差异越大, 其信度越高, 测量个体内差异越小, 其信度越高; 最后, 测量信度越高,

其检测统计效应所需样本量越小。

那么, 到底什么原因使得个体测量的信度如此重要呢? 从理论上讲, 结合上述 3 条测量效度的统计定律, 加上在实际应用和实践中, 对于测量的效度是无法直接进行测量的(否则就没必要进行研究了, 因为疾病或特质有效性已经解决的话, 就意味着相应科学问题已经回答), 从实际应用上讲, 教育实践中的“因材施教”和临床实践中的“精准医疗”都体现了个体差异测量研究的价值。高信度测量意味着更易于区别不同的学生或病人, 而在不同场合的测量稳定性也更好, 综合来看, 高信度测量对个体差异研究和应用转化至关重要。

近 10 年来, 神经影像因其安全性和高时空精度的优势, 已经积累了大量数据, 成千上万的人脑影像已经上线并公开, 涵盖了人类在不同发展阶段和各类脑疾病障碍上的影像<sup>[2]</sup>, 由此而催生了开放式神经科学的出现, 推动了大型脑科学(比如人脑功能和脑疾病生物标记物)研究。个体差异研究的基础是统计力度, 其决定了检测实验效应的能力。大样本量是提高统计力度的因素之一, 然而如果测量信度不够, 就会产生对大样本量的不必要的需求。在此评论文章中, 研究团队采用蒙特卡洛方法对信度、样本量和效应量之间的关系进行了数值模拟。结果揭示: 在神经影像领域, 测量的信度局限将极大地增加研究对样本量的需求。神经影像测验的信度研究表明: 现有数据集中较少有足够的个体数据能获得高度可靠的脑连接测量。各国推出的各类大型脑计划中, 个体差异的基本和转化研究(教育和临床)是中国脑计划的核心和特色,

# Measurement Reliability: Mapping Its Anatomy



The Reliability of Clinical Diagnoses: State of the Art

Helena Chmura Kraemer

Department of Psychiatry and Behavioral Sciences, Stanford University (Emerita), Palo Alto, California 94301; and Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania 15213; email: hckhome@pacbell.net

Developing and Validating Clinical Questionnaires

Anthony J. Rosellini and Timothy A. Brown

Center for Anxiety and Related Disorders, Department of Psychological and Brain Sciences, Boston University, Boston, Massachusetts 02215, USA; email: ajrosell@bu.edu, t\_abrown@bu.edu



Annual Review of  
Clinical Psychology

## Keywords

validity, disorder, design, kappa

## Abstract

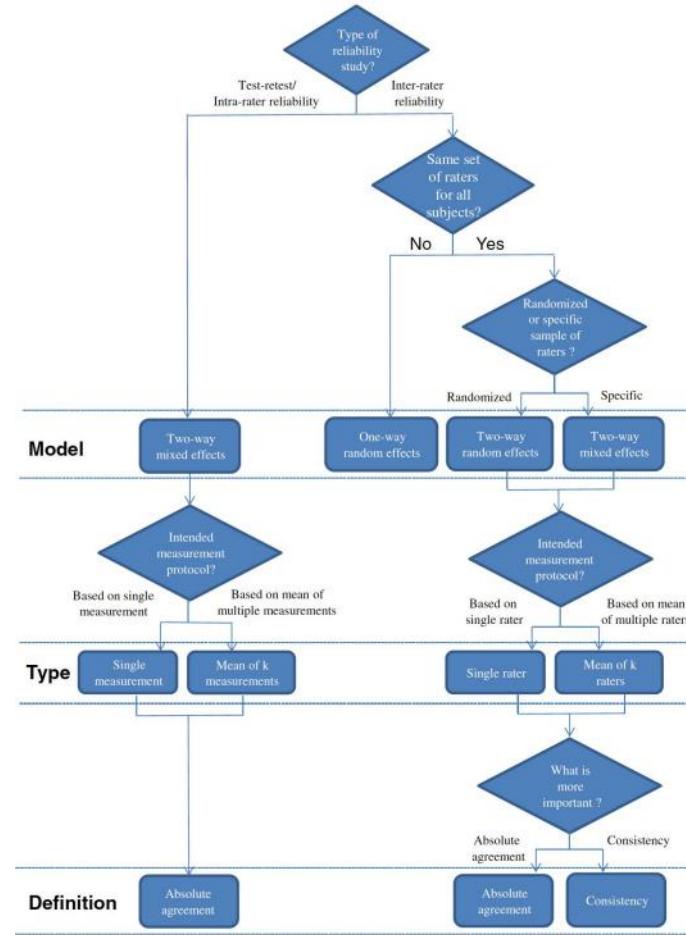
Reliability of clinical diagnosis is essential for good clinical decision making as well as productive clinical research. The current review emphasizes the distinction between a disorder and a diagnosis and between validity and reliability of diagnoses, and the relationships that exist between them. What is crucial is that reliable diagnoses are essential to establishing valid diagnoses. The present review discusses the theoretical background underlying the evaluation of diagnoses, possible designs of reliability studies, estimation of the reliability coefficient, the standards for assessment of reliability, and strategies for improving reliability without compromising validity.

# Measurement Reliability: Statistical Assessment

## *Analysis of Variance Models Used in Developing Intraclass Correlation Coefficient Definitions*

Case label	Model	Assumptions
Case 1: One-way random effects	$x_{ij} = \mu + r_i + w_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$ .	$\mu$ (the population mean for all observations) is constant; $r_i$ (the row effects) are random, independent, and normally distributed with mean 0 and variance $\sigma_r^2$ ; and $w_{ij}$ (residual effects) are random, independent, and normally distributed with mean 0 and variance $\sigma_w^2$ . Moreover, the effects $r_i$ and $w_{ij}$ are pairwise independent.
Case 2: Two-way random effects, with interaction	$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$ .	$\mu$ and $r_i$ are as before; $c_j$ (the column effects) are random, independent, and normally distributed with mean 0 and variance $\sigma_c^2$ ; $rc_{ij}$ (the interaction effects) are random, independent, and normally distributed with mean 0 and $\sigma_{rc}^2$ ; and $e_{ij}$ (residual effects) are random, independent, and normally distributed with mean 0 and variance $\sigma_e^2$ . Moreover, all the effects are pairwise independent.
Case 2A: Two-way random effects, interaction absent	$x_{ij} = \mu + r_i + c_j + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$ .	Same as for Case 2 except that there is no interaction effect.
Case 3: Two-way mixed effect model, with interaction	$x_{ij} = \mu + r_i + c_j + rc_{ij} + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$ .	Same as for Case 2 except that $c_j$ are fixed so that $\sum c_j = 0$ , $\sum_{j=1}^k rc_{ij} = 0$ , and the parameter corresponding to $\sigma_r^2$ in Case 2 is $\theta_r^2 = \sum c_j^2/(k-1)$ .
Case 3A: Two-way mixed model, interaction absent	$x_{ij} = \mu + r_i + c_j + e_{ij}$ where $i = 1, \dots, n$ and $j = 1, \dots, k$ .	Same as for Case 3 except that there is no interaction effect.

		true variance true variance + error variance
Shrout and Fleiss (1979)	Formulas for Calculating ICC <sup>b</sup>	
Convention <sup>b</sup>		
ICC (1,1)		$\frac{MS_R - MS_W}{MS_R + (k+1)MS_W}$
—		$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$
ICC (2,1)		$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$
ICC (3,1)		$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$
—		$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$
ICC (1,k)		$\frac{MS_R - MS_W}{MS_R}$
—		$\frac{MS_R - MS_E}{MS_R}$
ICC (2,k)		$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$
ICC (3,k)		$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$
—		$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$



# Measurement Reliability: Statistical Properties

## Harnessing reliability for neuroscience research

Neuroscientists are amassing the large-scale datasets needed to study individual differences and identify biomarkers. However, measurement reliability within individual samples is often suboptimal, thereby requiring unnecessarily large samples. We focus our comment on reliability in neuroimaging and provide examples of how the reliability can be increased.

Xi-Nan Zuo, Ting Xu and Michael Peter Milham

The neuroimaging community has made significant strides towards collecting large-scale neuroimaging datasets, which – until the past decade – had seemed out of reach. Between initiatives focused on the aggregation and open sharing of previously collected datasets and de novo data generation initiatives tasked with the creation of community resources, tens of thousands of datasets are now available, covering a wide range of developmental statuses and disorders, and many more will soon be available. Such open data are allowing researchers to increase the scale of their studies, to apply various learning strategies (for example, artificial intelligence) with ambitions of brain-based biomarker discovery and to address questions regarding the reproducibility of findings, all at a pace that is unprecedented in imaging. However, based on the findings of recent works<sup>1–3</sup>, few of the datasets generated to date contain enough data points to achieve highly reliable measures of brain connectivity. Although our examination of this critical deficiency focuses on the field of neuroimaging, the implications of our argument and the statistical principles discussed are broadly applicable.

**Scoping the problem**  
Our concern is simpler: researchers working hard to amass large-scale datasets are often doing so using uncoordinated data-generation initiatives, but failing to optimize their data collections for relevant reliabilities (for example, test-retest, between raters, etc.<sup>4</sup>). They may be collecting larger amounts of suboptimal data, rather than smaller amounts of higher-quality data, a trade-off that does not bode well for the field, particularly when it comes to making inferences and predictions at the individual level. We believe that this mistake can be avoided by critical assessments of reliability upfront.

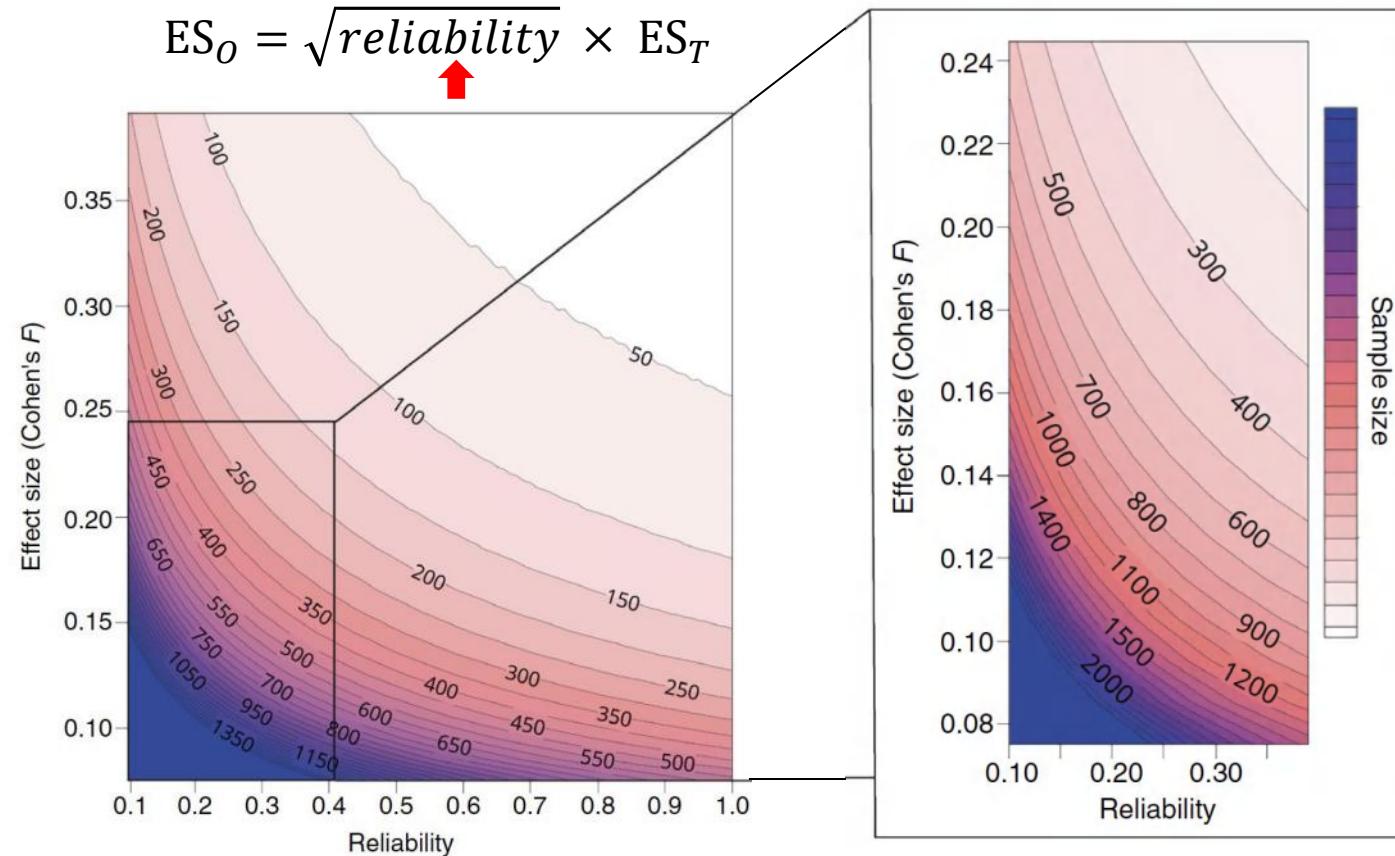
The trade-off we observe occurring in neuroimaging reflects a general tendency in neuroscience. Statistical power is

fundamental to studies of individual differences, as it determines our ability to detect effects of interest. While sample size is readily recognized as a key determinant of statistical power, measurement reliabilities are less commonly considered and at best are only indirectly considered when estimating required sample sizes. This is unfortunate, as statistical theory dictates that reliability places an upper limit on the number of observations required to detect differences in the presence of noise.

The interplay between reliability, sample size and effect size in determinants of statistical power is commonly underappreciated in the field. To facilitate a direct discussion of these factors, Fig. 1 depicts the impact of measurement reliability and effect size on the sample sizes required to achieve desirable levels of statistical power (for example, 80%). These relations are not heavily dependent on the specific items of statistical inference employed (for example, two-sample t-tests or one-way analysis of variance (ANOVA)). Estimates were generated using the power package in R and are highly congruent with results from Monte Carlo simulations<sup>5</sup>. With respect to neuroscience, where the bulk of findings report effect sizes ranging from modest to moderate<sup>6</sup>, the figure makes obvious our point that increasing reliability can dramatically reduce the sample size requirements (and therefore cost) for achieving statistically appropriate designs.

In neuroscience, one of the measures employed in experimental neuroimaging is reliability, which is typically underappreciated in neuroscience research. Whether one is focusing on imaging, electrophysiology, neuroinflammatory markers, microbiomes, cognitive neuroscience paradigms or on-person devices, it is essential that we consider measurement reliability and its determinants.

For MRI-based neuroimaging, a repeated theme across the various modalities (for example, diffusion,



"True" correlation( $x, y$ ) =  $\frac{\text{Sample correlation}(x, y)}{\sqrt{\text{Reliability}(x) \cdot \text{Reliability}(y)}}$

$\rightarrow r(\text{measure } A, \text{ measure } B) = r(\text{true } A, \text{ true } B) \sqrt{\text{reliability } (\text{Measure } A) \text{reliability } (\text{Measure } B)}$

# Measurement Reliability: Comparative Methods

## Fisher's $r$ to $z$ Transformation for ICCs

Because textbooks generally give only the formula for converting interclass  $rs$  to  $z'$ , it is important to note that the formula is different for converting ICCs, which Fisher also designates as  $r$  (Fisher, 1938, p. 225). Rather than using the interclass formula

$$z' = \frac{1}{2} \log \frac{1+r}{1-r},$$

one uses the formula

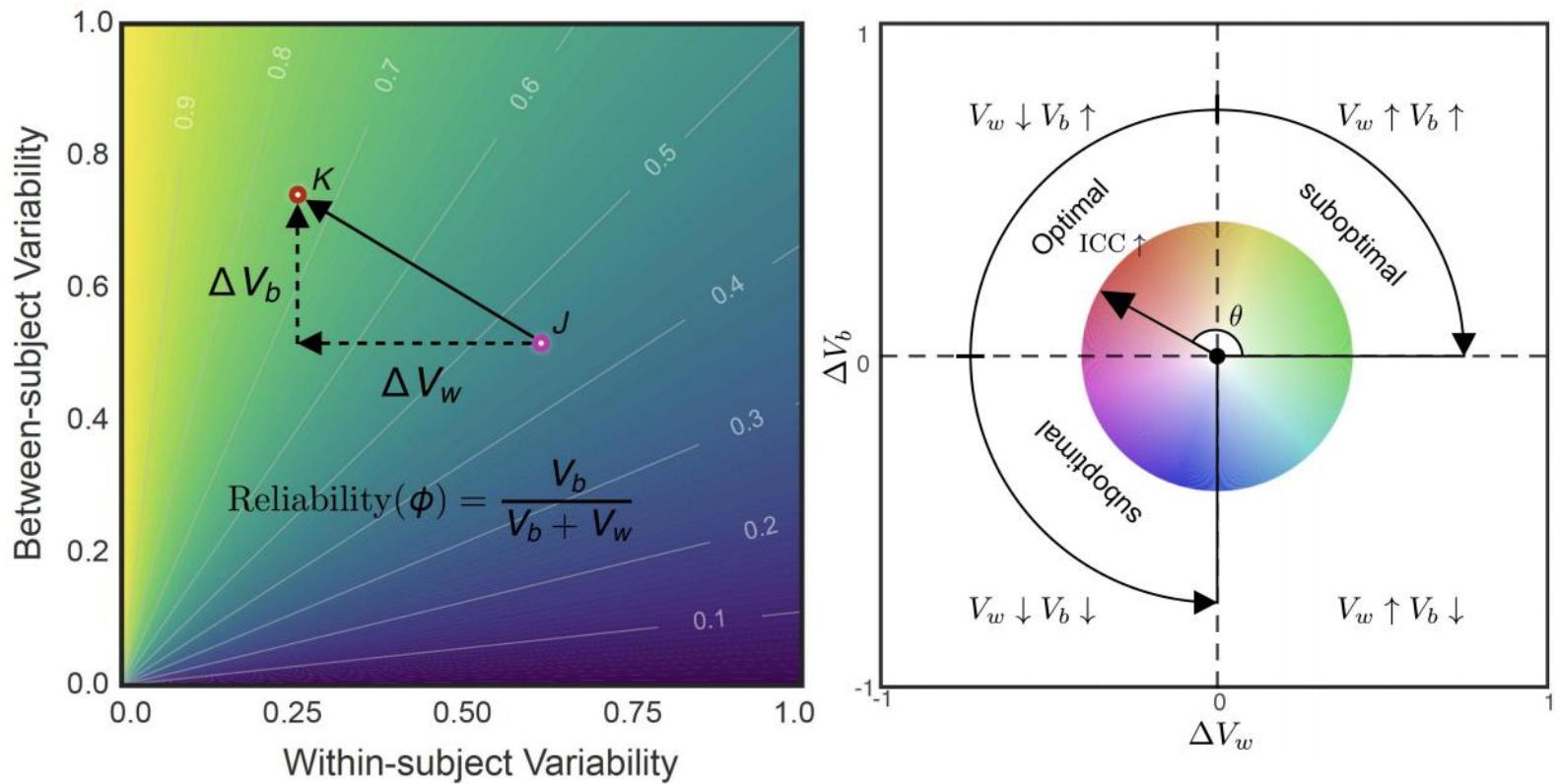
$$z_i = \frac{1}{2} \log \frac{1 + (k-1)r}{1 - r},$$

where  $k$  is the number of observations made on each object of measurement. The variance of the above statistic is

$$\sigma^2 = \frac{k}{2(n-2)(k-1)},$$

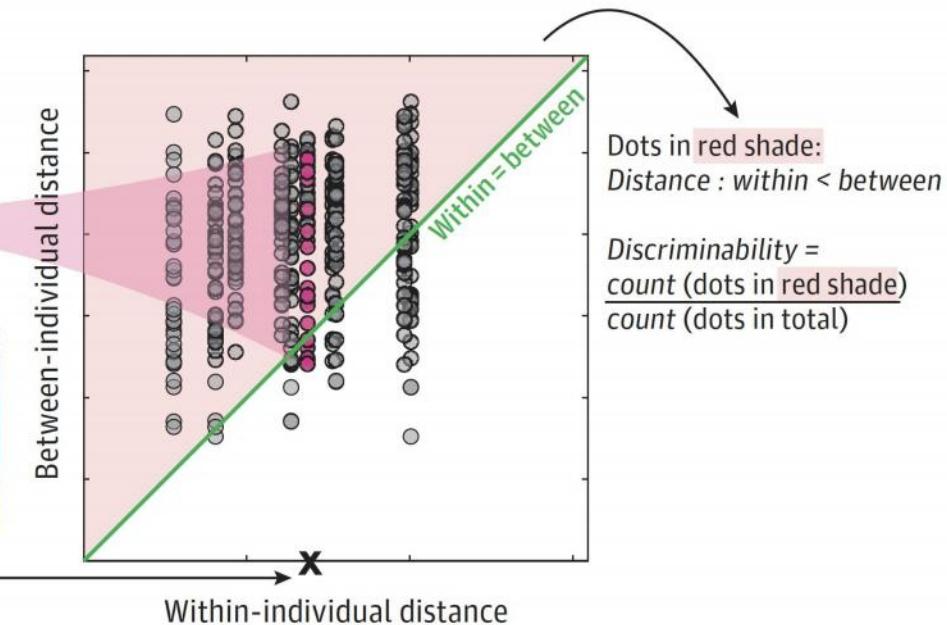
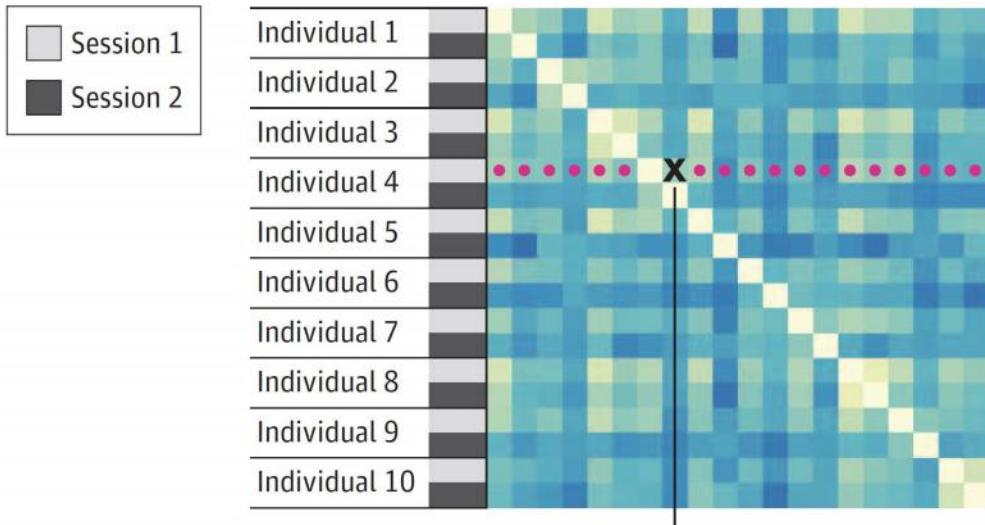
where  $n$  is the number of objects of measurement and  $k$  is, again, the number of observations.

$$\lambda = \frac{z_1 - z_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \sim N(0,1)$$



# Measurement Reliability: Statistical Extension

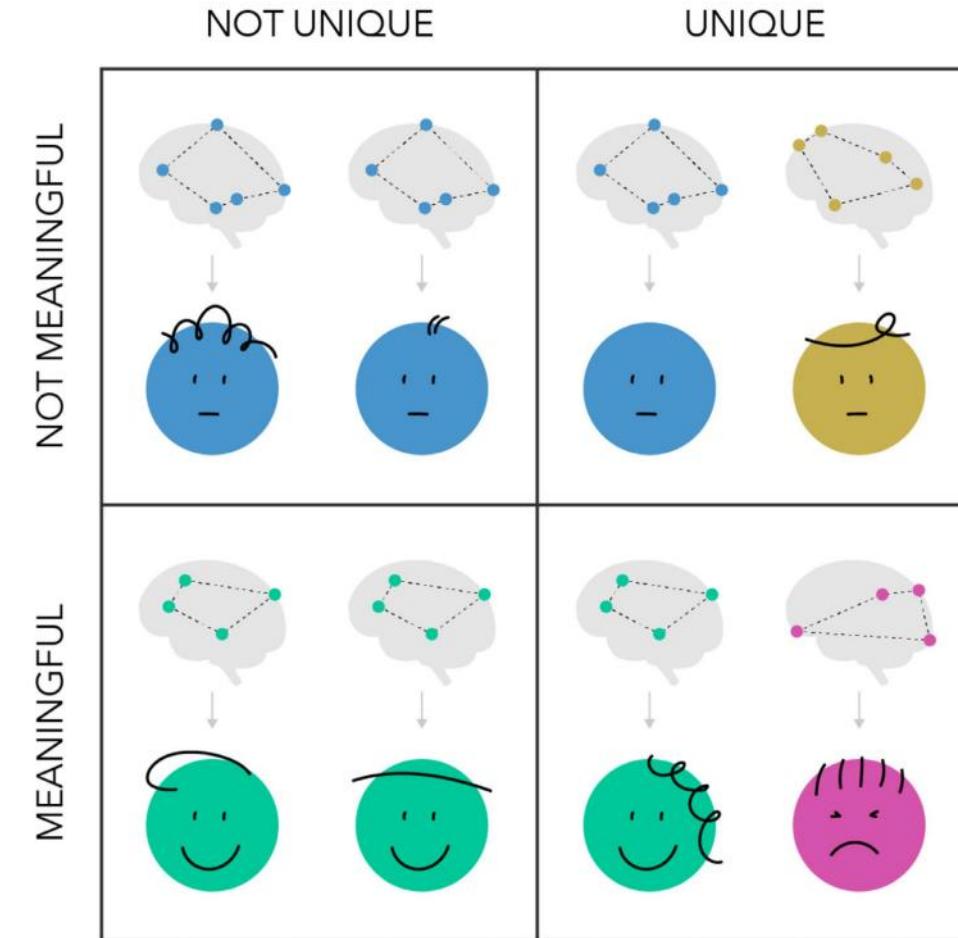
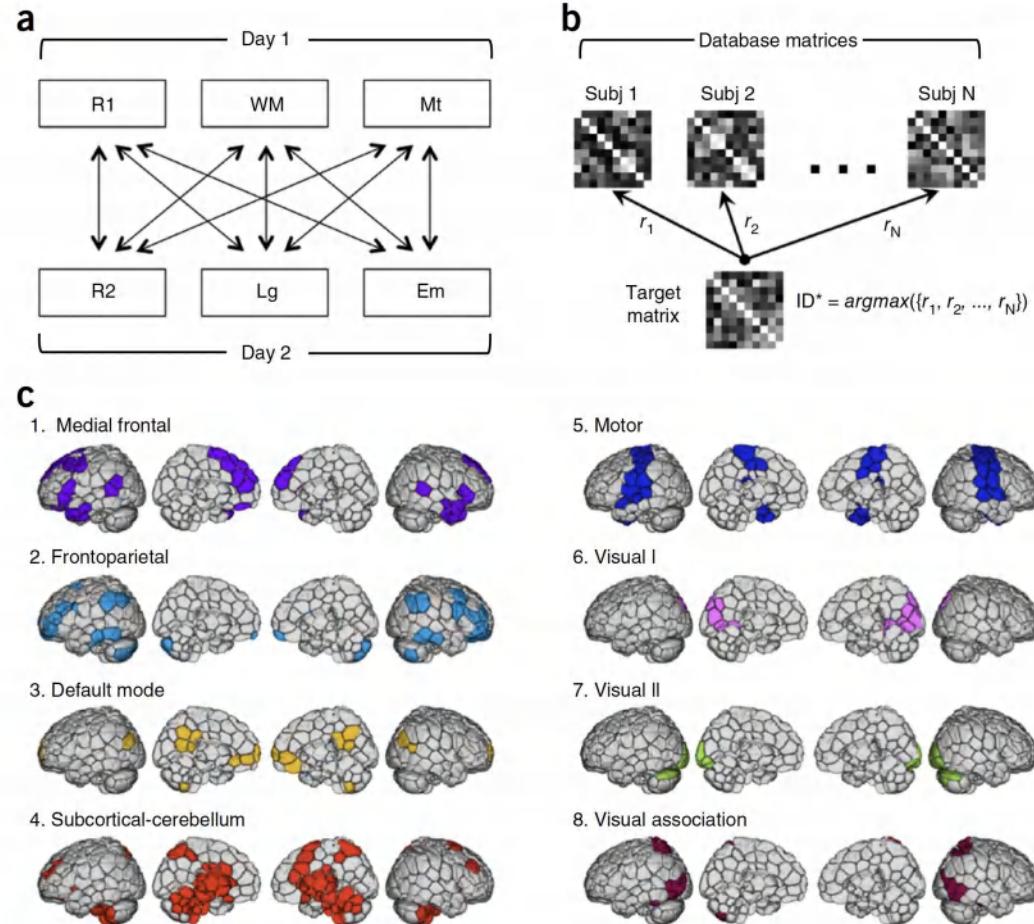
A Within- and between-individuals distance matrix



$$D = 1 - \frac{\arctan\left(\frac{\sqrt{\sigma^2(3\sigma^2+4\sigma_\mu^2)}}{\sigma_\mu^2}\right)}{\pi} = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{\text{ICC}}{\sqrt{(1-\text{ICC})(\text{ICC}+3)}}\right)$$

# Measurement Reliability:

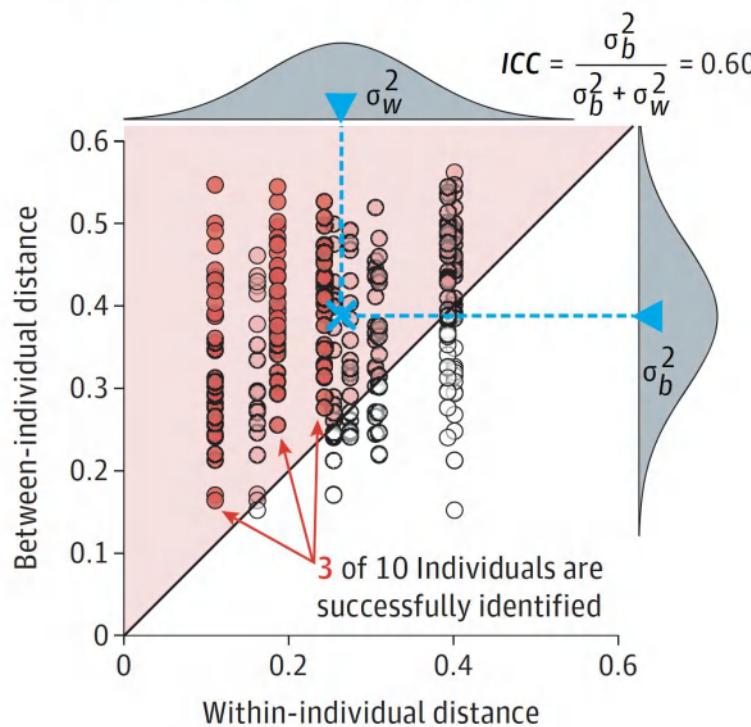
$$F_{index} = \rho D + (1 - \rho)D^{n-1}$$



# Measurement Reliability: Achieve Clinical Utility

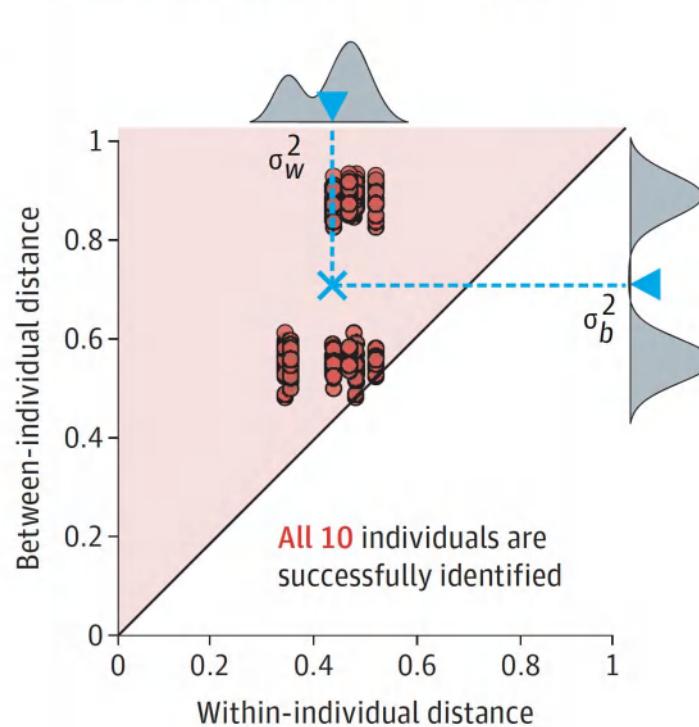
**B** Repeatability metrics (Discriminability, 0.84)

$$\text{Identification rate} = \frac{\text{count (red columns)}}{\text{count (individuals)}} = 0.30$$



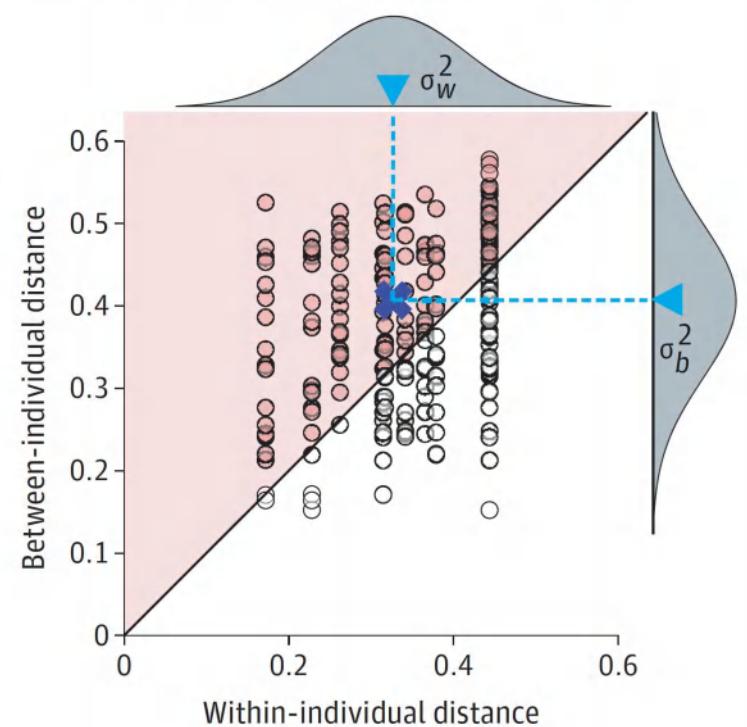
**C** The individual can be identified but the gaussian assumption is violated

Discriminability, 1.00; ICC, 0.62; FP, 1.00



**D** Relatively reliable data but the identification rate is 0

Discriminability, 0.77; ICC, 0.57; FP, 0.00



# Measurement Reliability: A Guidance on Optimal Use



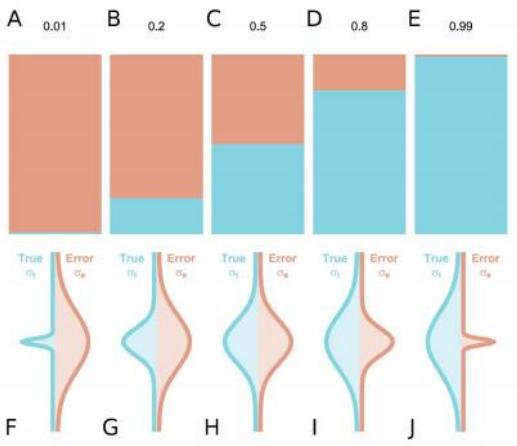
We need to talk about reliability: making better use of test-retest studies for study design and interpretation

Granville J. Matheson

Department of Clinical Neuroscience, Center for Psychiatry Research, Karolinska Institutet and Stockholm County Council, Stockholm, Sweden

## ABSTRACT

Neuroimaging, in addition to many other fields of clinical research, is both time-consuming and expensive, and recruitable patients can be scarce. These constraints limit the possibility of large-sample experimental designs, and often lead to statistically underpowered studies. This problem is exacerbated by the use of outcome measures whose accuracy is sometimes insufficient to answer the scientific questions posed. Reliability is usually assessed in validation studies using healthy participants, however these results are often not easily applicable to clinical studies examining different populations. I present a new method and tools for using summary statistics from previously published test-retest studies to approximate the reliability of outcomes in new samples. In this way, the feasibility of a new study can be assessed during planning stages, and before collecting any new data. An R package called *relifeas* also accompanies this article for performing these calculations. In summary, these methods and tools will allow researchers to avoid performing costly studies which are, by virtue of their design, unlikely to yield informative conclusions.



Submitted 27 September 2018  
Accepted 7 April 2019  
Published 24 May 2019

Corresponding author  
Granville J. Matheson,  
granville.matheson@gmail.com

Academic editor  
Andrew Gray

Additional Information and  
Declarations can be found on  
page 20

DOI:10.7717/peerj.6918

Copyright  
2019 Matheson

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

## PLOS COMPUTATIONAL BIOLOGY

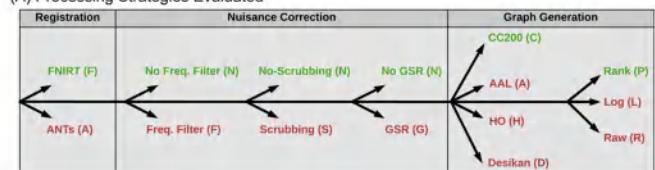
### RESEARCH ARTICLE

Eliminating accidental deviations to minimize generalization error and maximize replicability: Applications in connectomics and genomics

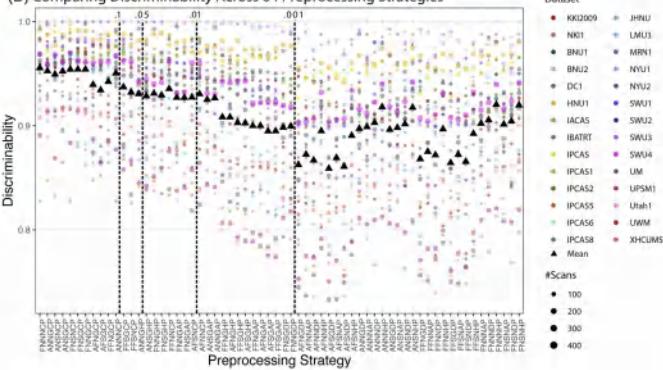
Eric W. Bridgeford<sup>1</sup>, Shangsi Wang<sup>2</sup>, Zeyi Wang<sup>2</sup>, Ting Xu<sup>3</sup>, Cameron Craddock<sup>2,3</sup>, Jayanta Dey<sup>4</sup>, Gregory Kiar<sup>5</sup>, William Gray-Rock<sup>6</sup>, Carlo Colantoni<sup>7</sup>, Christopher Douville<sup>8</sup>, Stephanie Noble<sup>9</sup>, Carey E. Priebe<sup>10</sup>, Brian Caffo<sup>11</sup>, Michael Bernstein<sup>12</sup>, Xi-Nian Zuo<sup>13</sup>, Consortium for Reliability and Reproducibility, Joshua T. Vogelstein<sup>1,2\*</sup>

<sup>1</sup> Johns Hopkins University, Baltimore, Maryland, United States of America, <sup>2</sup> McGill University, Montreal, Quebec, Canada, <sup>3</sup> Child Mind Institute, New York, New York, United States of America, <sup>4</sup> Yale University, New Haven, Connecticut, United States of America, <sup>5</sup> State Key Laboratory of Cognitive Neuroscience and Learning, International Data Group-McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, <sup>6</sup> Progressive Learning, Baltimore, Maryland, United States of America

### (A) Processing Strategies Evaluated



### (B) Comparing Discriminability Across 64 Preprocessing Strategies



## Building Functional Network Neuroscience for Reliable Individual Differences

Chao Jiang<sup>1,2,3</sup>, Richard Betzel<sup>4</sup>, Ye He<sup>5</sup>, Yin-Shan Wang<sup>1,6,7</sup>, Xiu-Xia Xing<sup>3,5,6</sup>, and Xi-Nian Zuo<sup>1,3,6,7,8</sup>

<sup>1</sup> State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

<sup>2</sup> School of Psychology, Capital Normal University, Beijing 100045, China

<sup>3</sup> Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

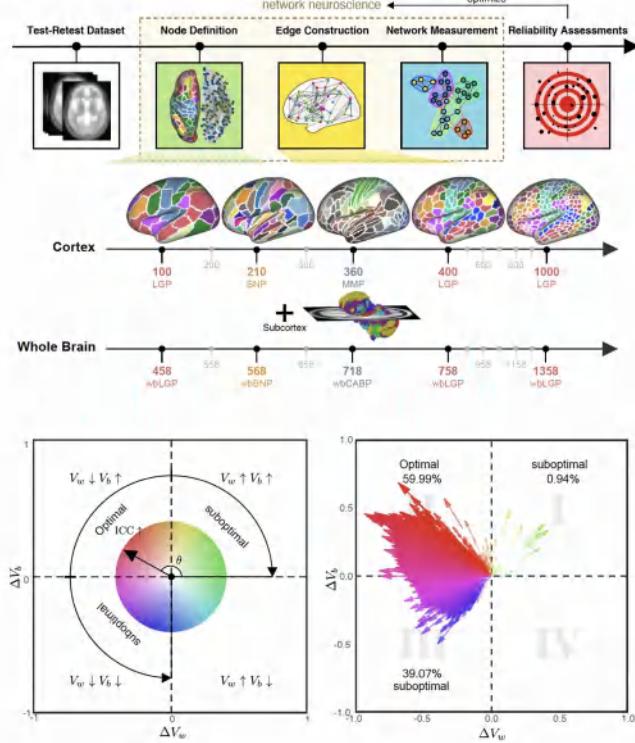
<sup>4</sup> Department of Psychology, University of Southern California, Los Angeles, California, United States

<sup>5</sup> School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

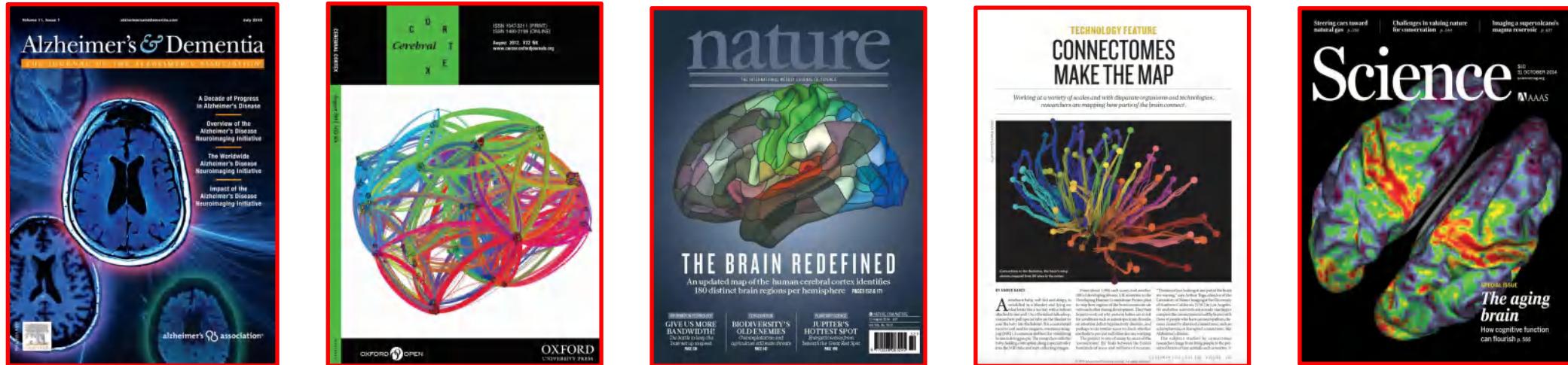
<sup>6</sup> National Basic Science Data Center, Beijing 100190, China

<sup>7</sup> McDonnell Institute for Brain Research, Beijing Normal University, Beijing 100875, China

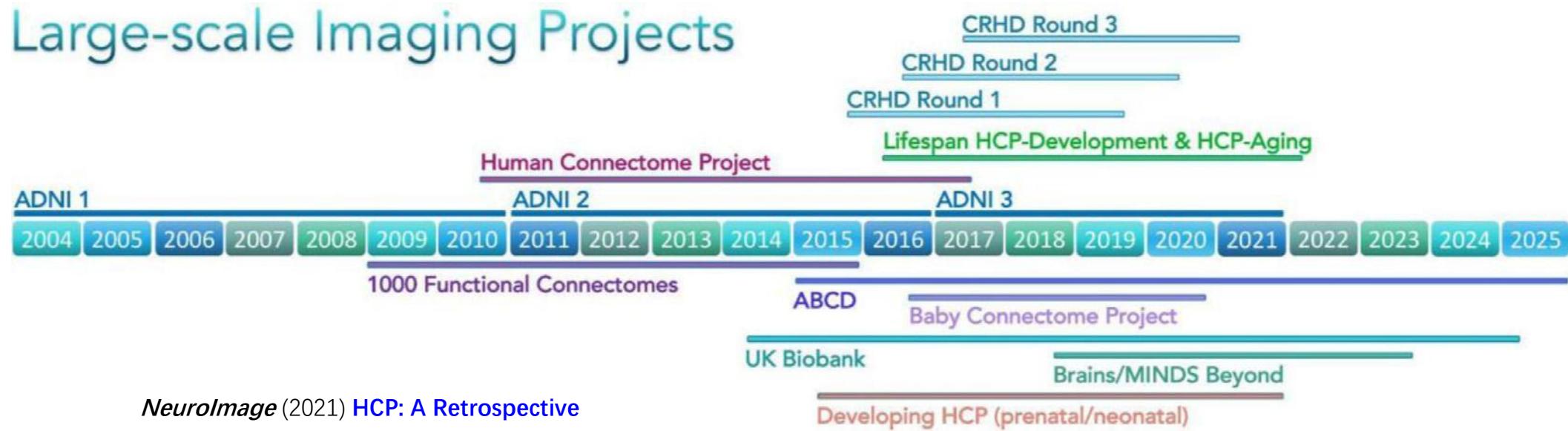
<sup>8</sup> Department of Applied Mathematics, College of Mathematics, Beijing University of Technology, Beijing 100124, China



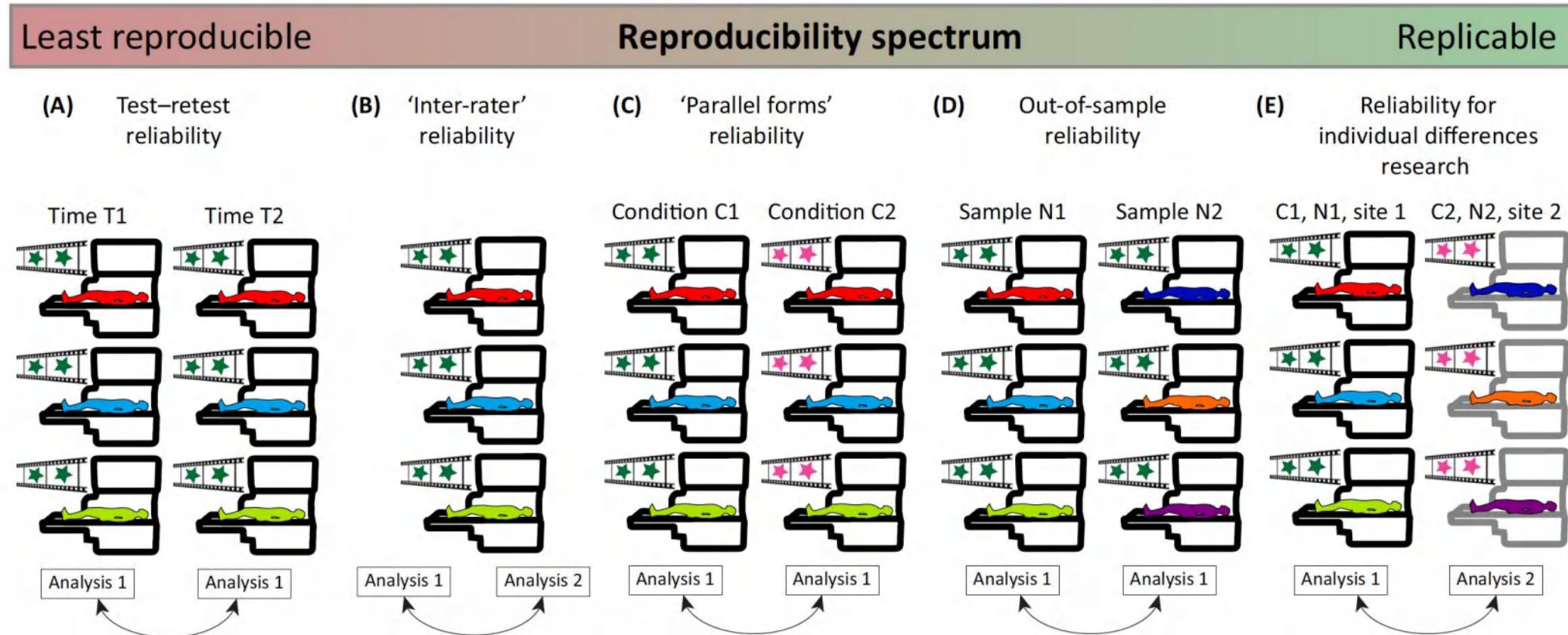
# Measurement Reliability: Human Brain Mapping



## Large-scale Imaging Projects

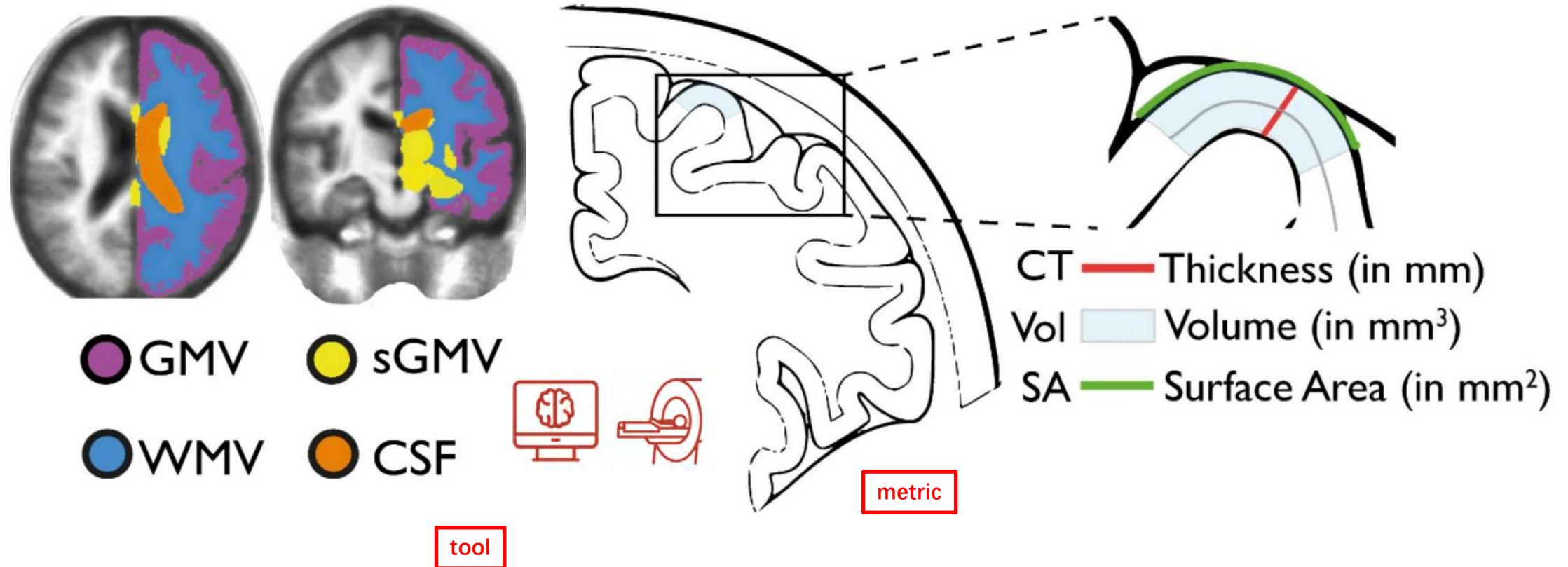


# Measurement Reliability: Human Brain Mapping

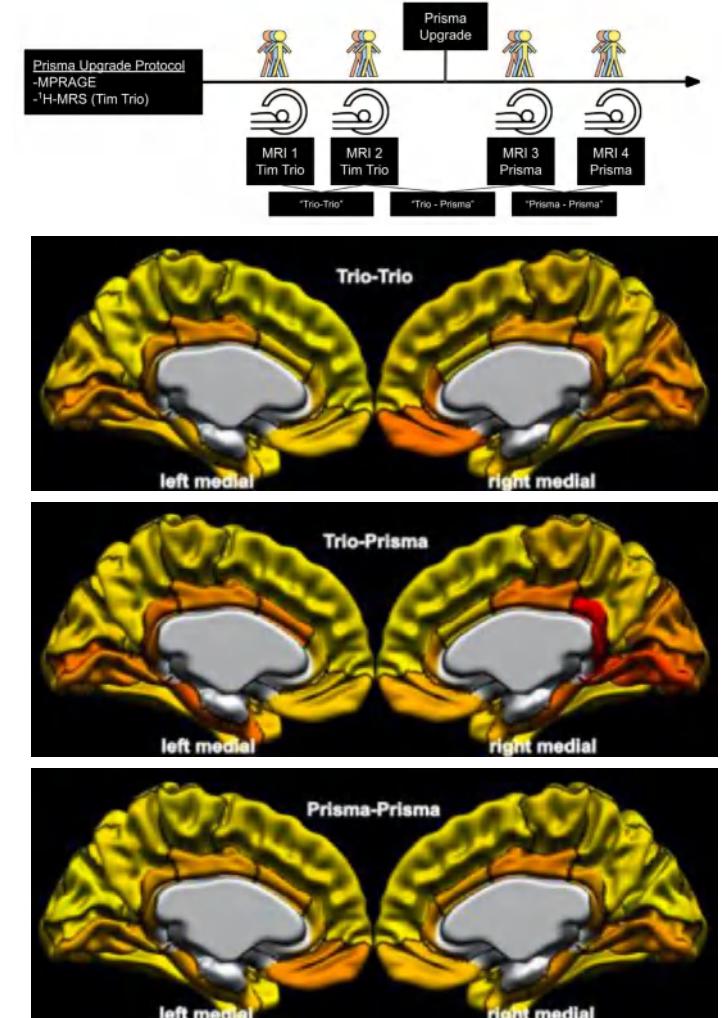
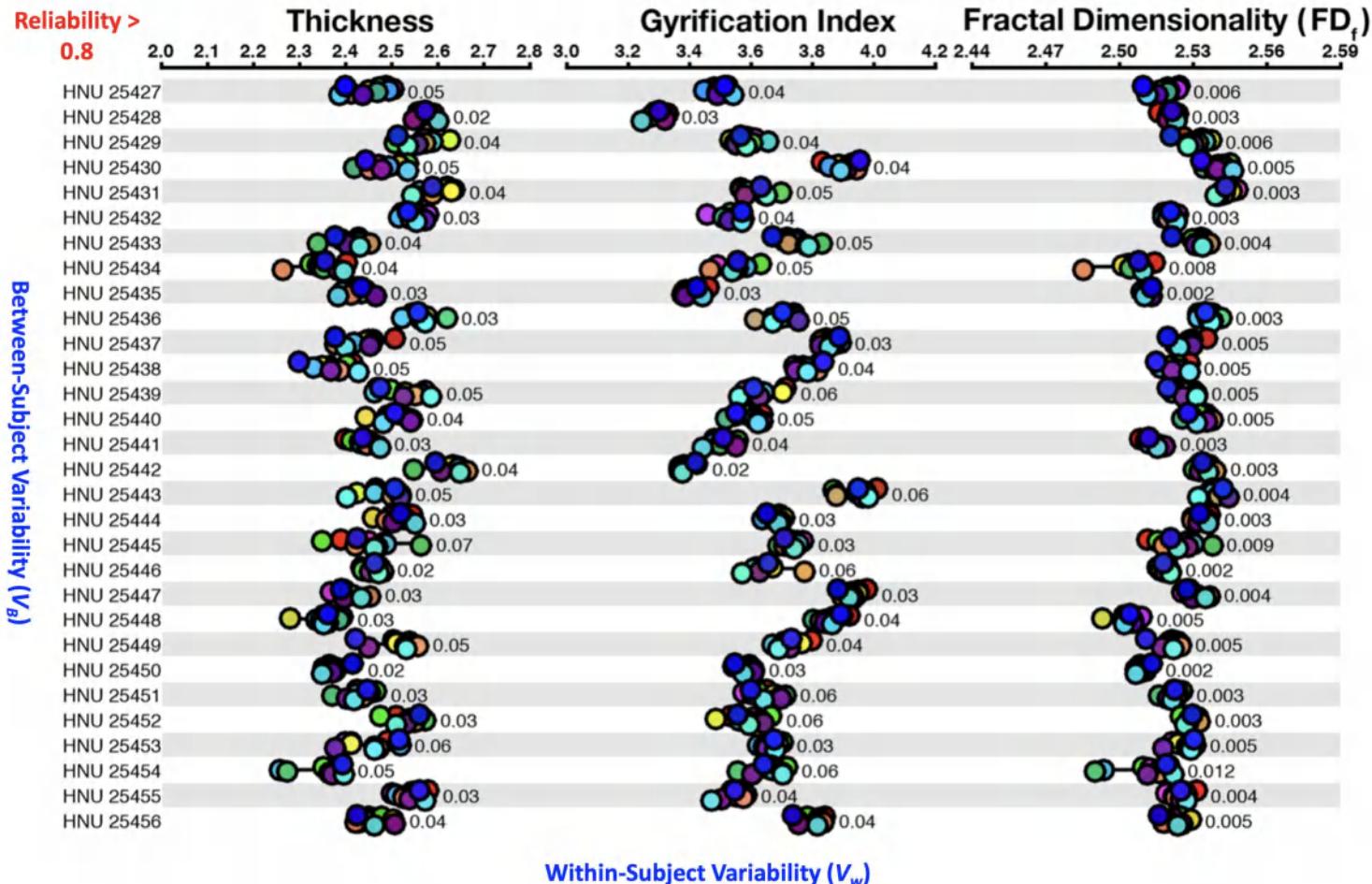


# Measurement Reliability: Brain Morphology

target

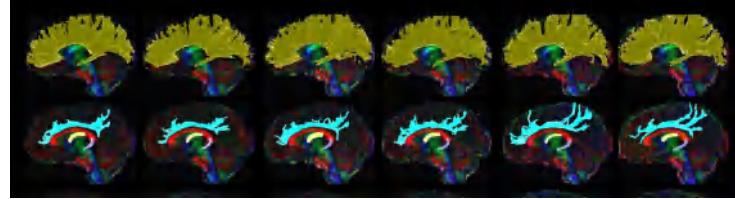


# Measurement Reliability: Brain Morphology



*Brain Informatics* (2017); *Brain Informatics* (2019); *NeuroImage* (2021)

# Measurement Reliability: Brain Microstructure



Measurements	Intrasession ICC		
	DTI30-2	DTI30-1	DTI15-2
<i>Corpus callosum</i>			
FA	<b>0.97</b>	<b>0.90</b>	<b>0.89</b>
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>0.94</b>	<b>0.92</b>	<b>0.90</b>
FC	<b>0.95</b>	<b>0.90</b>	<b>0.93</b>
ML (mm)	<b>0.93</b>	<b>0.93</b>	<b>0.99</b>
TV (voxels/l)	<b>0.95</b>	<b>0.92</b>	<b>0.96</b>
FD (FC/voxel)	<b>0.86</b>	<b>0.84</b>	<b>0.97</b>

Measurements	Intersession ICC		
	DTI30-2	DTI30-1	DTI15-2
<i>Corpus callosum</i>			
FA	<b>0.90</b>	<b>0.88</b>	<b>0.86</b>
MD ( $\mu\text{m}^2/\text{ms}$ )	<b>0.90</b>	<b>0.76</b>	<b>0.89</b>
FC	<b>0.91</b>	<b>0.93</b>	0.39
ML (mm)	<b>0.92</b>	<b>0.93</b>	0.57
TV (voxels/l)	<b>0.84</b>	<b>0.95</b>	0.26
FD (FC/voxel)	<b>0.91</b>	<b>0.92</b>	0.62

TABLE 1. UNWEIGHTED METRICS, 82-NODE CONNECTOME

Intrasite	Session 1	Session 2	CV%	ICC
	Avg. $\pm$ SD	Avg. $\pm$ SD		
K	12.8 $\pm$ 1.4	12.4 $\pm$ 1.4	3.21	0.89
L	2.26 $\pm$ 0.13	2.28 $\pm$ 0.12	1.44	0.92
C	0.54 $\pm$ 0.02	0.53 $\pm$ 0.02	3.40	0.51
B	0.032 $\pm$ 0.003	0.033 $\pm$ 0.003	2.73	0.90
E	0.51 $\pm$ 0.02	0.51 $\pm$ 0.02	1.18	0.92
E <sub>loc</sub>	0.75 $\pm$ 0.01	0.75 $\pm$ 0.02	1.79	0.54

Intersite	Site 1	Site 2	CV%	ICC
	Avg. $\pm$ SD	Avg. $\pm$ SD		
K	13.8 $\pm$ 1.6	12.7 $\pm$ 2.6	6.26	0.79
L	2.22 $\pm$ 0.13	2.30 $\pm$ 0.21	2.26	0.82
C	0.59 $\pm$ 0.01	0.55 $\pm$ 0.03	3.05	0.59
B	0.031 $\pm$ 0.003	0.033 $\pm$ 0.005	4.00	0.82
E	0.52 $\pm$ 0.03	0.50 $\pm$ 0.04	1.94	0.84
E <sub>loc</sub>	0.78 $\pm$ 0.01	0.76 $\pm$ 0.02	1.61	0.63

metric



tool

TABLE 2. WEIGHED METRICS, 82-NODE CONNECTOME

Intrasite	Session 1	Session 2	CV%	ICC
	Avg. $\pm$ SD	Avg. $\pm$ SD		
K <sup>w</sup>	6600 $\pm$ 950	6400 $\pm$ 840	4.31	0.84
L <sup>w</sup>	0.083 $\pm$ 0.05	0.096 $\pm$ 0.08	34.75	0.67
C <sup>w</sup>	15.9 $\pm$ 3.4	15.1 $\pm$ 3.0	5.50	0.83
B <sup>w</sup>	0.067 $\pm$ 0.005	0.069 $\pm$ 0.004	4.58	0.52

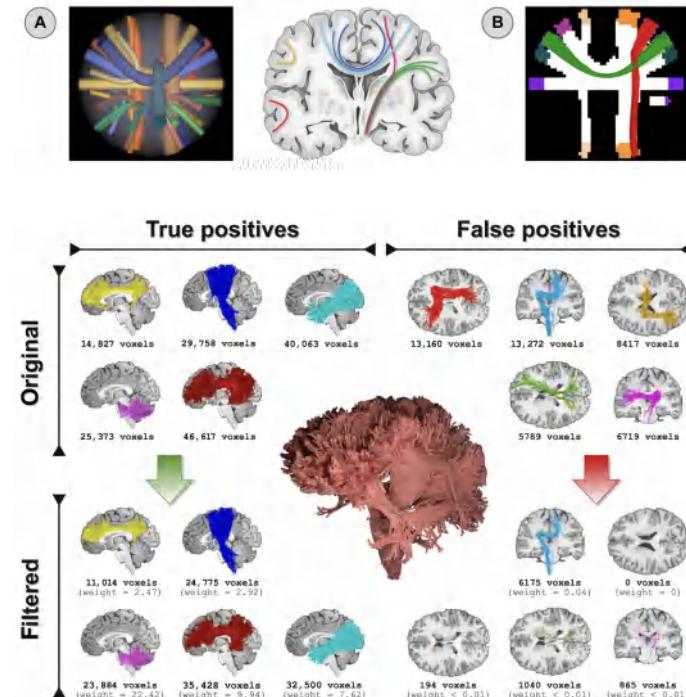
  

Intersite	Site 1	Site 2	CV%	ICC
	Avg. $\pm$ SD	Avg. $\pm$ SD		
K <sup>w</sup>	7500 $\pm$ 970	7000 $\pm$ 1700	10.17	0.69
L <sup>w</sup>	0.190 $\pm$ 0.12	0.170 $\pm$ 0.15	14.46	0.94
C <sup>w</sup>	18.1 $\pm$ 3.7	16.4 $\pm$ 5.7	10.57	0.81
B <sup>w</sup>	0.066 $\pm$ 0.003	0.067 $\pm$ 0.006	2.63	0.70

## NEUROSCIENCE

A new method for accurate *in vivo* mapping of human brain connections using microstructural and anatomical information

Simona Schiavi<sup>1,2</sup>, Mario Ocampo-Pineda<sup>1</sup>, Muhamed Barakovic<sup>2</sup>, Laurent Petit<sup>3</sup>, Maxime Descoteaux<sup>4</sup>, Jean-Philippe Thiran<sup>2,5</sup>, Alessandro Daducci<sup>1\*</sup>



# Measurement Reliability: Human Brain Function

Trends in Cognitive Sciences

CellPress

## Feature Review

### Building a Science of Individual Differences from fMRI

Julien Dubois<sup>1,\*</sup> and Ralph Adolphs<sup>1</sup>

To date, fMRI research has been concerned primarily with evincing generic principles of brain function through averaging data from multiple subjects. Given rapid developments in both hardware and analysis tools, the field is now poised to study fMRI-derived measures in individual subjects, and to relate these to psychological traits or genetic variations. We discuss issues of validity, reliability and statistical assessment that arise when the focus shifts to individual subjects and that are applicable also to other imaging modalities. We emphasize that individual assessment of neural function with fMRI presents specific challenges and necessitates careful consideration of anatomical and vascular between-subject variability as well as sources of within-subject variability.

#### From the Group to the Individual

Brain imaging with blood oxygen level-dependent functional magnetic resonance imaging (BOLD fMRI) has been used extensively since the early 1990s to understand generic aspects of brain function, typically by averaging data across individuals to improve the signal-to-noise ratio (SNR). The statistical benefits of averaging across subjects have also been leveraged in group comparisons, for example in studies of clinical populations. However, these studies have historically fallen short of a proper characterization of brain function at the level of the individual. While the importance of a fully personalized investigation of brain function has been recognized for several years [1,2], only recent technological advances now make it possible. For example, there are advances already at the acquisition level, such as higher field strength and faster acquisition, which have led to substantial SNR improvements [3]. Attempts at interpreting individual subject fMRI measurements have become a major focus in the past 5 years or so, partly driven by the rise of resting-state fMRI (Box 1). There is interest in examining individual differences in relation to healthy aging [4,5], personality [6], intelligence [7,8], mood [9] and genetic polymorphism [10]. On the clinical side, there are considerable efforts to use fMRI to classify individual subjects as patient or control [11,12; reviewed in [13,14]], to select treatment [15], or predict future outcome [16]; reviewed in [17].

Several issues arise when the focus shifts from group averaging to the comparison of the statistics of individual subjects. The issues can be framed in terms of key concepts from behavioral research on individual differences, namely validity and reliability. Validity asks whether the individual differences we measure with BOLD fMRI really reflect what we intend to measure. One specific concern for validity is whether we are indeed comparing functionally homologous regions across subjects. Another is whether we are indeed comparing neural function because BOLD fMRI only provides an indirect measure of neural activity. Reliability, by contrast, asks whether a finding is stable in the face of variations that should not matter. Reliability is notably hindered by relatively well-understood noise sources such as motion and subject physiology, as well as by less well-understood ones such as neuro- or vasoactive substances that we might not

**Trends**  
Interpretation of fMRI data at the level of individual brains is essential for characterizing brain function, health and disease.  
  
Two core challenges are validity (do we measure what we intend to measure?) and reliability (are our measures stable in the face of variations that should not matter?) of fMRI-derived individual differences; these challenges can be partly addressed with recent tools.

Interpretation of single-subject fMRI measures relies on establishing a relationship between independent measures in the same subjects. Out-of-sample prediction should be used over cross-validation analysis.

Accumulation of large samples through consortia and data sharing, as well as careful attention to statistical power issues, are crucial for reproducible research.

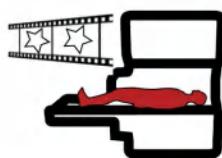
Whole-brain characterizations in naturalistic conditions, such as while watching a movie or listening to a story, may provide an alternative to resting-state data that permits a rich link to sensory and semantic stimulus variables.

<sup>1</sup>Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA

\*Correspondence:  
j.dubois@tumail.com (J. Dubois).

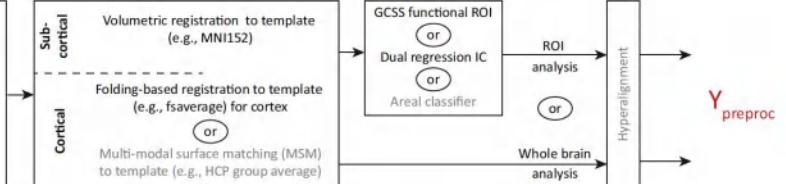
## Proposed Generic Analytical Pipeline for Individual Differences Research in fMRI

### (A) Acquire and preprocess fMRI data $Y_{\text{raw}}$



- [1. GDC if large gradient nonlinearity]
2. Realignment
- [3. Slice-timing correction if TR>2s]
4. Field inhomogeneity correction
5. Coregistration to T1w
6. Grand mean scaling
7. Temporal filtering

### (B) Map data to a common space



### (C) Derive fMRI statistic $s$

$$Y_{\text{preproc}} = \beta_{\text{sig}} X + \beta_{\text{nuis}} X + \epsilon$$

Model signal (if applicable)  
HRF + derivatives  
or other constrained basis set  
or  
FIR

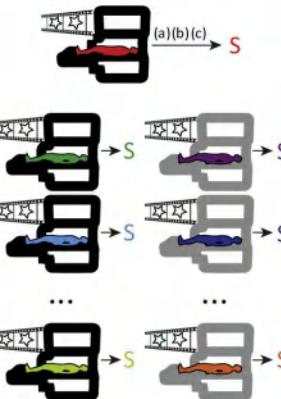
Motion regressors (12 or 24 or 36)  
and  
Scrubbing  
and/or  
Tissue regressors (WM, CSF, GS, CompCor)  
and/or  
Physiological regressors (RETROICor, RVHRCor)  
and/or  
Noise ICs (ICA-FIX, ME-ICA)

Model noise

Normalization

Statistical analysis  $\rightarrow s$

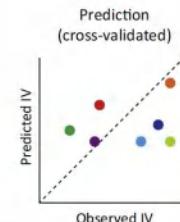
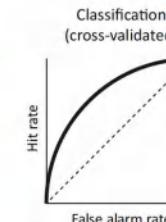
### (D) Repeat many times ( $n > 100$ )



### (E) Establish predictive value of $s$ for an Independent variable of interest $IV$

Full model:  $IV = f_{\text{FULL}} (s, \text{site}, \text{motion}, \text{other individual measures})$

Null model:  $IV = f_{\text{NULL}} (\text{site}, \text{motion}, \text{other individual measures})$



$AUC_{\text{FULL}} > AUC_{\text{NULL}}$ ?  
Permutation statistics

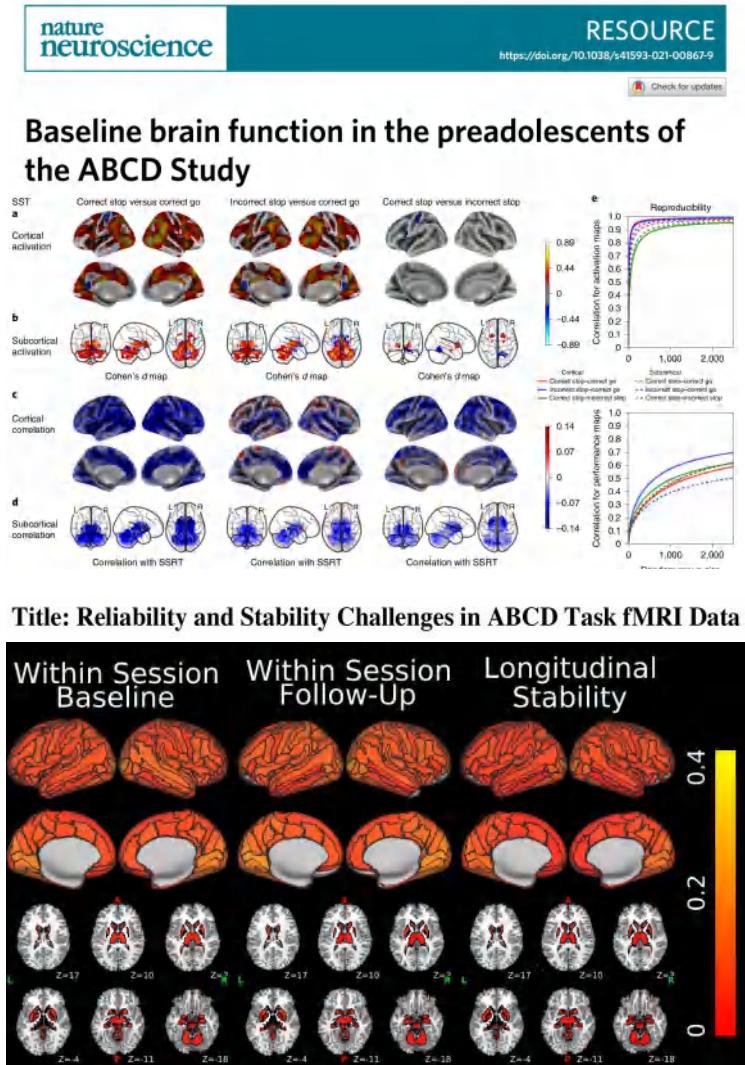
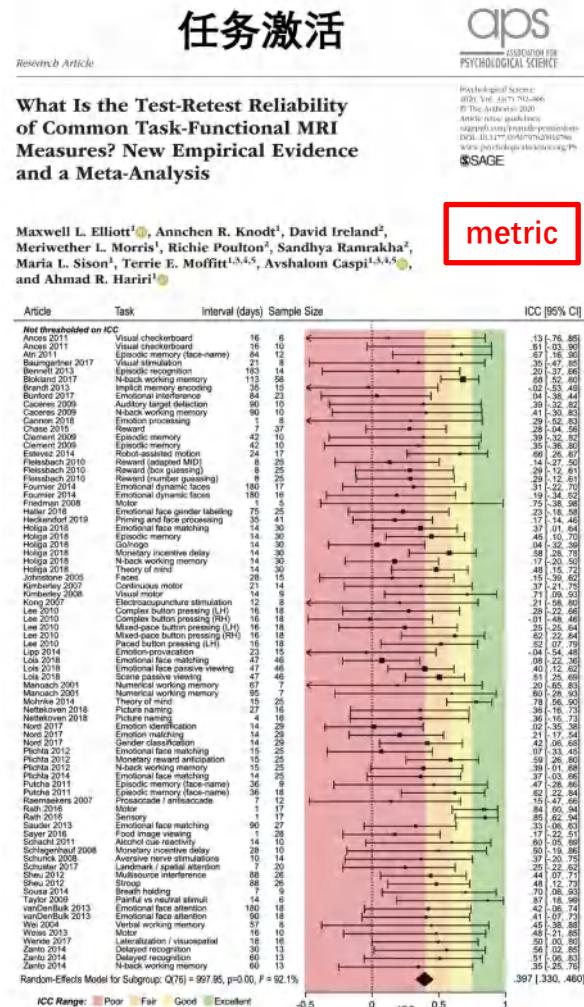
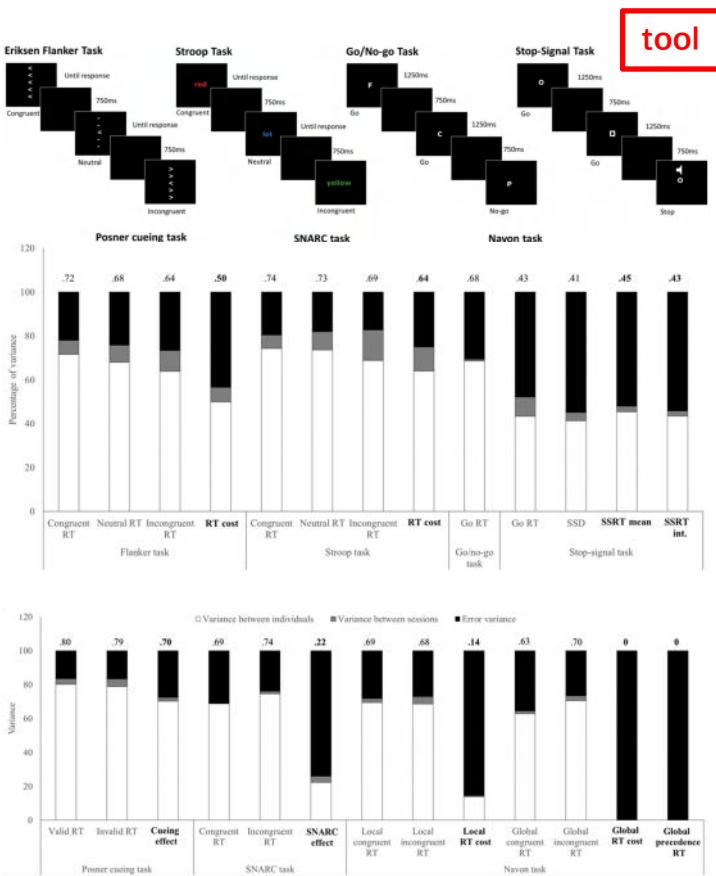
$R^2_{\text{FULL}} > R^2_{\text{NULL}}$ ?  
Permutation statistics

# Measurement Reliability: Task Brain Function

target

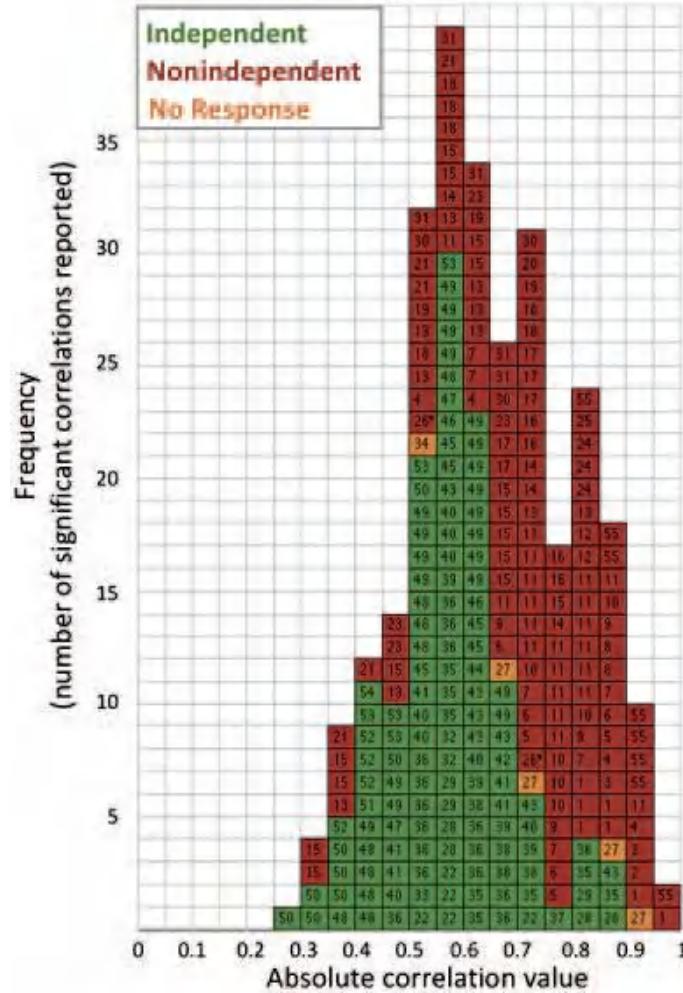
## The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences

Craig Hedge<sup>1</sup> · Georgina Powell<sup>1</sup> · Petroc Sumner<sup>1</sup>



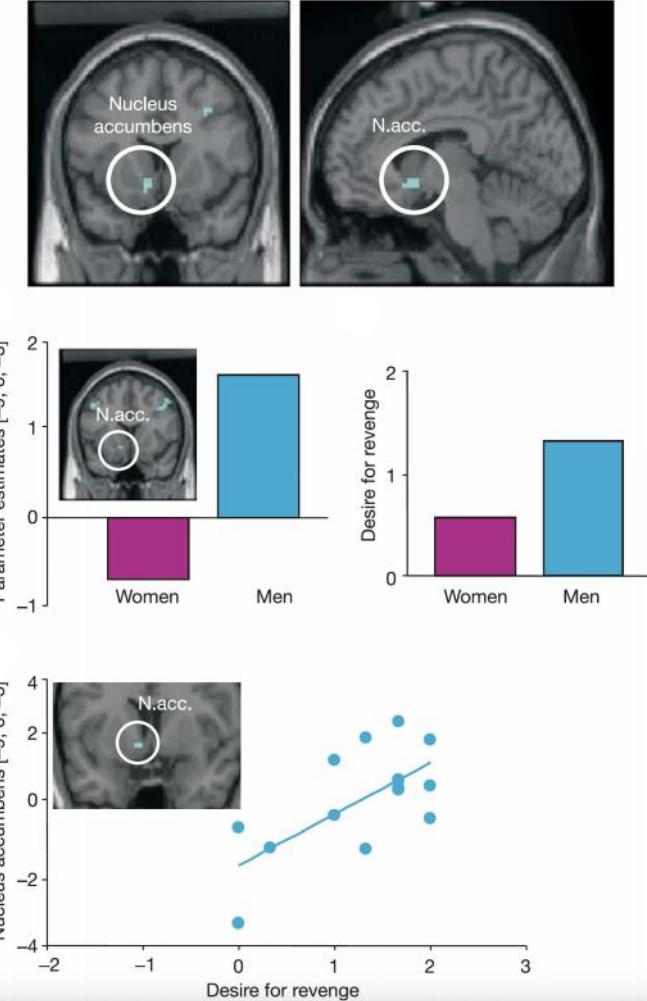
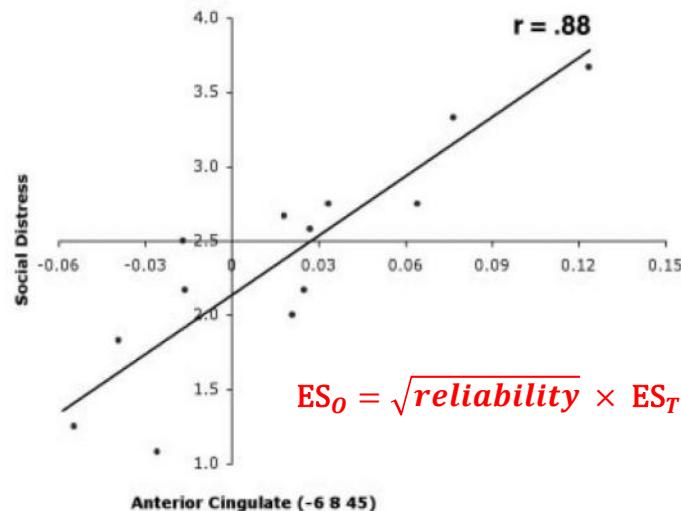
# Measurement Reliability: Task Brain Function

target



## Does Rejection Hurt? An fMRI Study of Social Exclusion

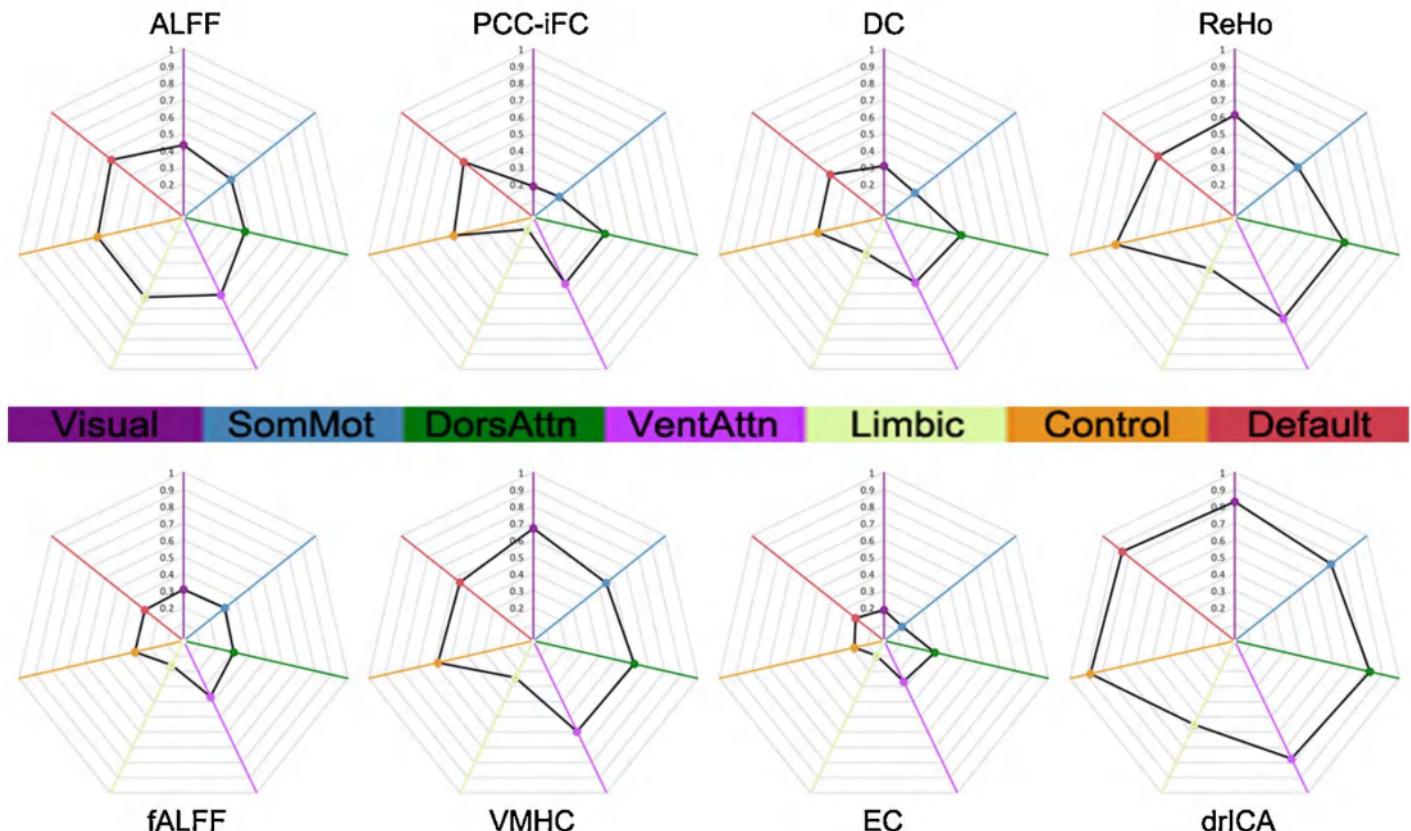
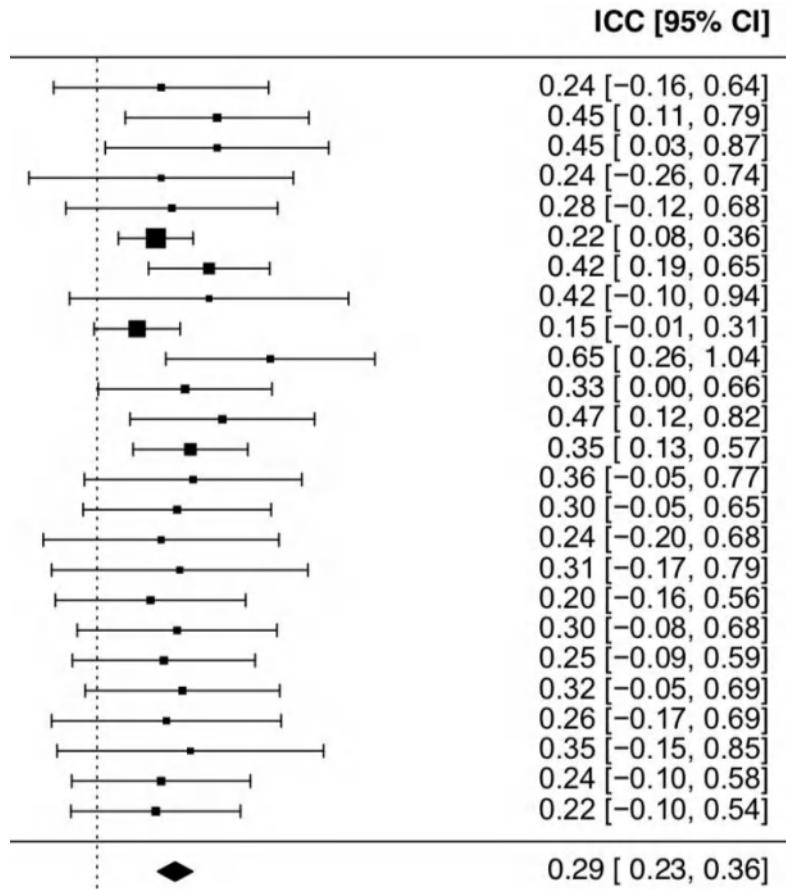
Naomi I. Eisenberger,<sup>1\*</sup> Matthew D. Lieberman,<sup>1</sup>  
Kipling D. Williams<sup>2</sup>



Perspectives on Psychological Science (2009); Science (2003); Nature (2006)

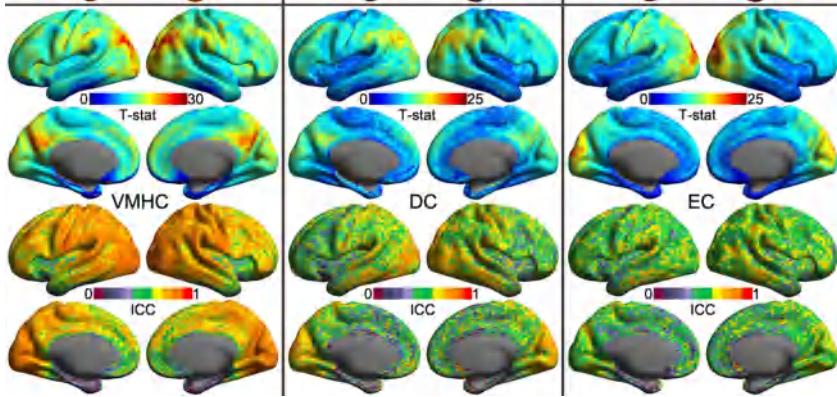
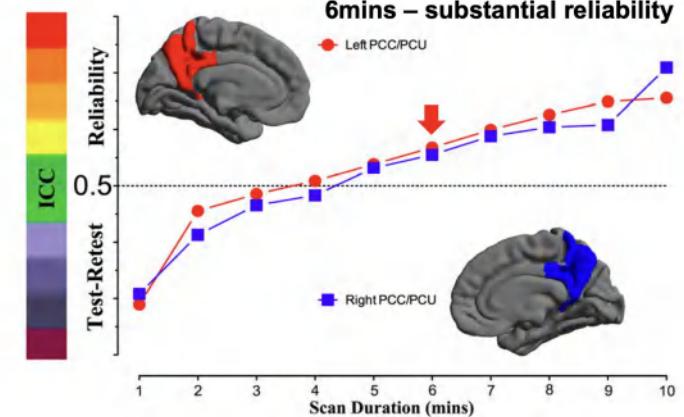
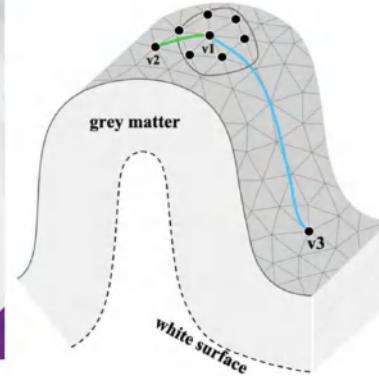
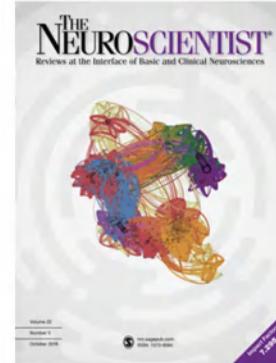
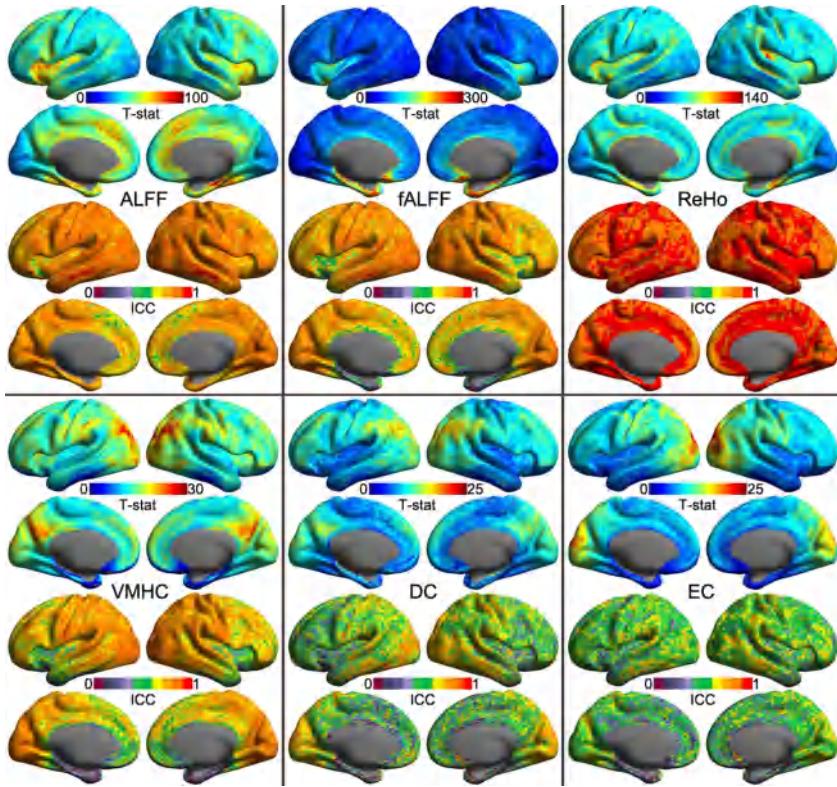
# Measurement Reliability: Rest Brain Function

target

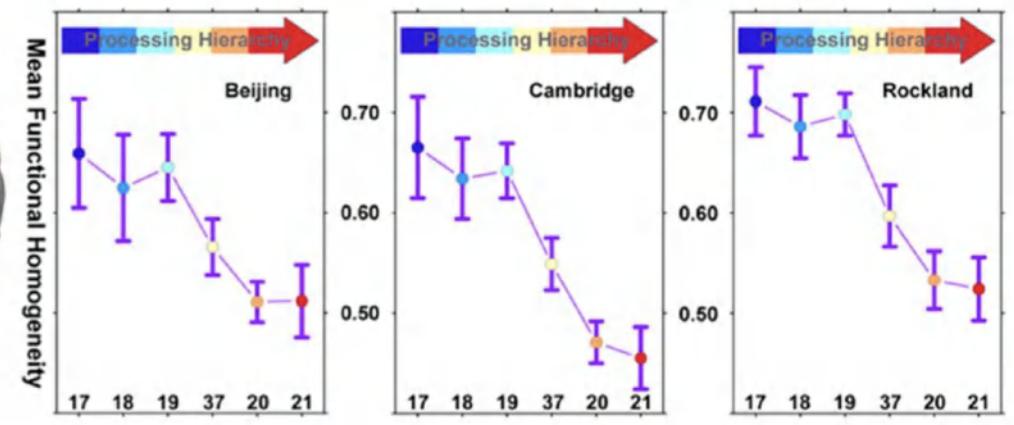
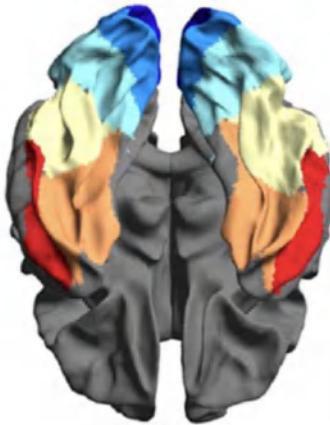


# Measurement Reliability: Rest Brain Function

target



Ventral Visual Stream – reproducibility & validity



# Measurement Reliability: Rest Brain Function

target

## Building Functional Network Neuroscience for Reliable Individual Differences

Chao Jiang<sup>1,2,3</sup>, Richard Betzel<sup>4</sup>, Ye He<sup>5</sup>, Yin-Shan Wang<sup>1,6,7</sup>, Xiu-Xia Xing<sup>8,9</sup>, and Xi-Nian Zuo<sup>1,3,6,7,9</sup>

<sup>1</sup>State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

<sup>2</sup>School of Psychology, Capital Normal University, Beijing 100048, China

<sup>3</sup>Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China

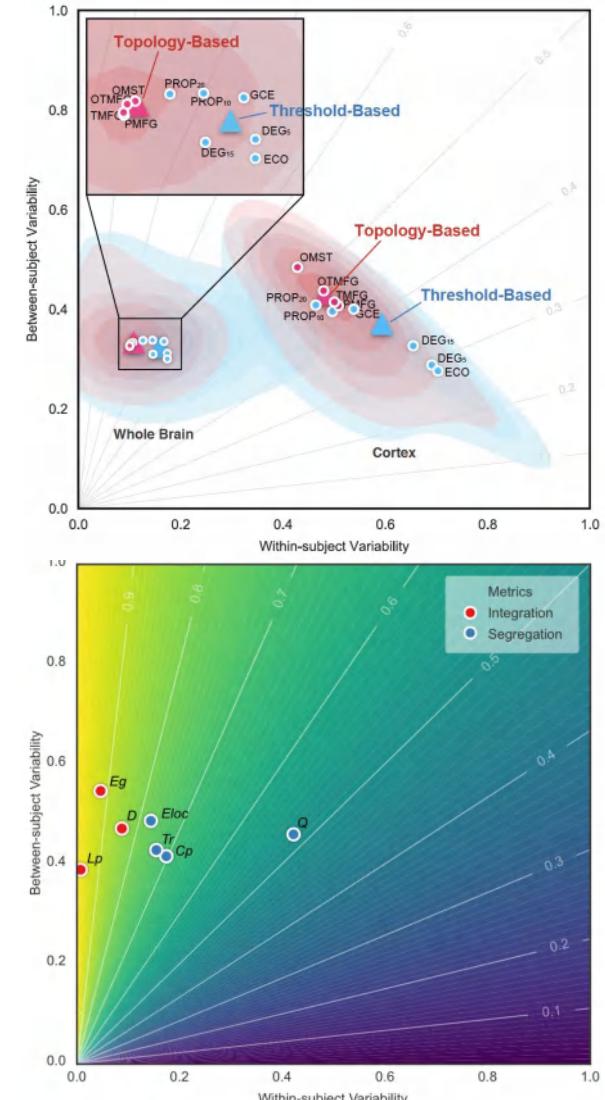
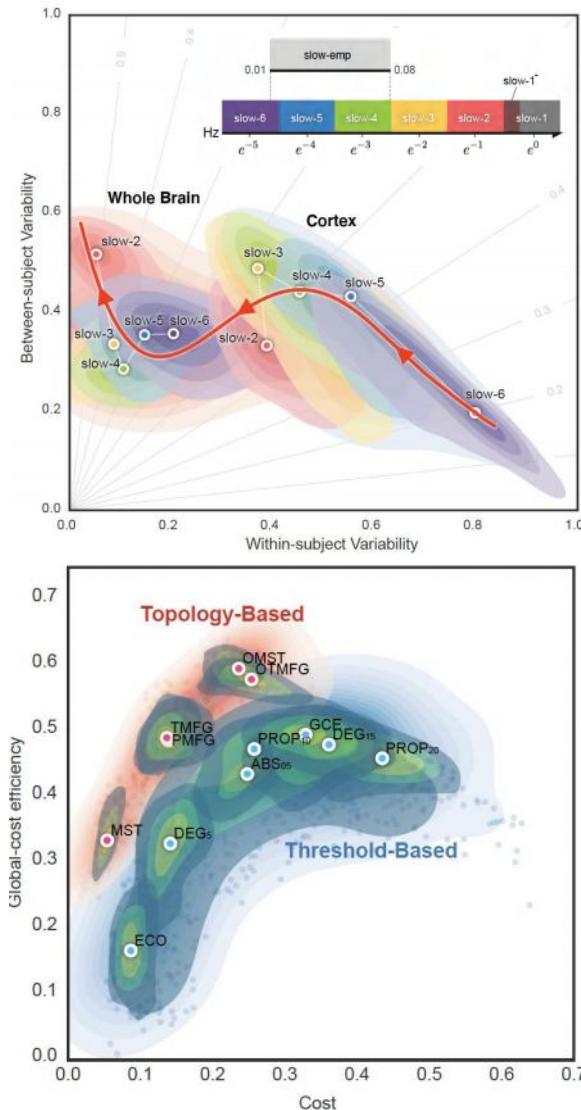
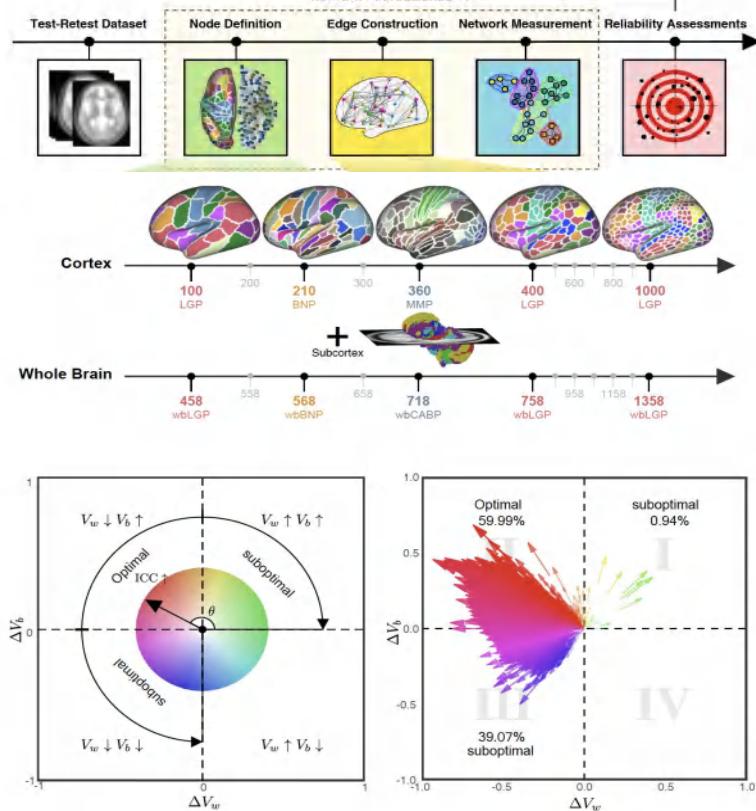
<sup>4</sup>Department of Psychology and Brain Sciences, Indiana University, Bloomington, United States

<sup>5</sup>School of Artificial Intelligence, Beijing Jiaotong University, Beijing 100044, China

<sup>6</sup>National Basic Science Data Center, Beijing 100190, China

<sup>7</sup>IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

<sup>8</sup>Department of Applied Mathematics, College of Mathematics, Faculty of Science, Beijing University of Technology, Beijing 100124, China



# Measurement Reliability: Moving to Translation

Trends in  
Cognitive Sciences

CellPress

Opinion

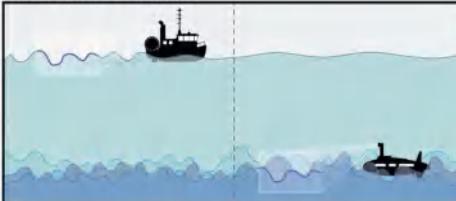
Is it time to put rest to rest?

Emily S. Finn<sup>1,2,9,\*</sup>

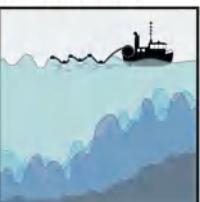
Annotated rest:



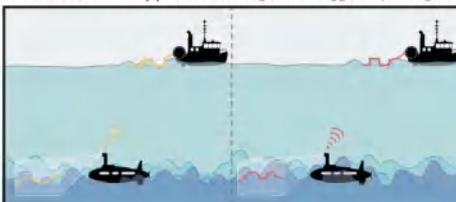
Task-signature echoes:



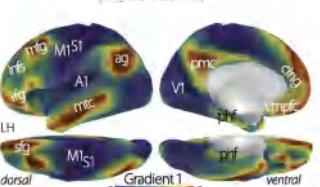
Naturalistic tasks:



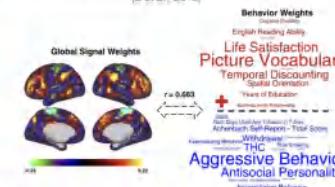
State-informed approaches: e.g., state-triggered paradigms



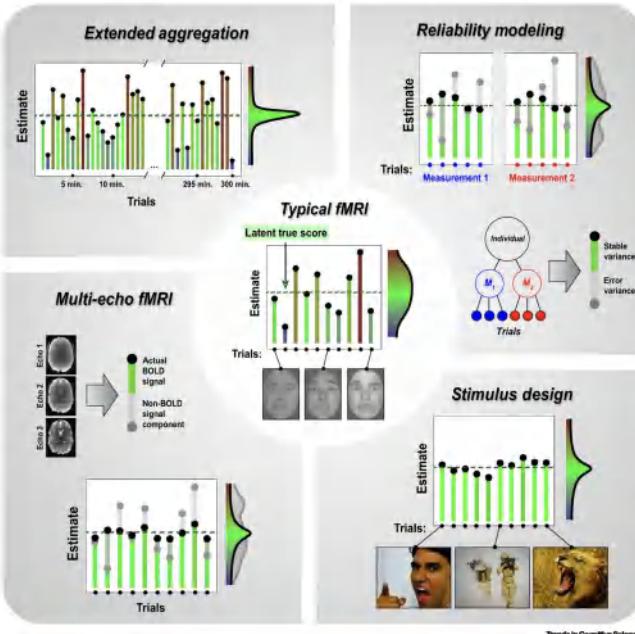
First gradient of default-mode network  
(Merello et al., 2018)



Regions associated with global signal intensity  
(Li et al., 2018)



Key figure  
Emerging strategies to generate more reliable fMRI measures



**高信度测量策略：扩展和聚合测量、精准建模稳定个体变异、移除无关生理干扰、设计新范式唤起个体差异。**  
**具体操作性推荐：**分段而多次fMRI扫描，获取总时长超过半小时的脑成像信号；针对特定脑功能反应提出数学模型（如潜变量）；采用ME-fMRI技术移除头动等生理干扰信号；设计并采用自然状态fMRI刺激范式。

Trends in  
Cognitive Sciences

CellPress

Review

Striving toward translation: strategies for reliable fMRI measurement

Maxwell L. Elliott,<sup>1,\*</sup> Annchen R. Knodt,<sup>1</sup> and Ahmad R. Hariri<sup>1</sup>

fMRI has considerable potential as a translational tool for understanding risk, prioritizing interventions, and improving the treatment of brain disorders. However, recent studies have found that many of the most widely used fMRI measures have low reliability, undermining this potential. Here, we argue that many fMRI measures are unreliable because they were designed to identify group effects, not to precisely quantify individual differences. We then highlight four emerging strategies [extended aggregation, reliability modeling, multi-echo fMRI (ME-fMRI), and stimulus design] that build on established psychometric properties to generate more precise and reliable fMRI measures. By adopting such strategies to improve reliability, we are optimistic that fMRI can fulfill its potential as a clinical tool.

**Highlights**

Since its introduction in 1992, fMRI has rapidly matured to become a powerful tool in neuroscience, allowing researchers to noninvasively map the functional organization of the average human brain and probe the brain bases of behaviors from the simple to the complex.

However, the translation of fMRI to clinical applications as well as the study of individual differences in fMRI has been limited to date. This limitation, in part, reflects the inadequate reliability of many of the most commonly used fMRI measures. Reliability is a prerequisite for the valid measurement of brain function in individuals.

We highlight four emerging strategies, each with roots in psychometrics, that have the potential to improve measurement reliability, thereby advancing the potential clinical utility of fMRI: (i) extended aggregation; (ii) reliability modeling; (iii) multi-echo fMRI; and (iv) stimulus design.

Cognitive neuroscience has revolutionized our understanding of how the brain supports behavioral functions ranging from basic sensory to complex cognitive processes. Based on these fundamental insights into brain and behavior, an emerging program of translational neuroscience seeks to identify individual differences in these patterns and, in so doing, inform the development of clinical biomarkers that can be used to predict disease risk, prioritize interventions, and improve treatment. Central to these efforts is fMRI, as it affords the noninvasive measurement of brain activity in behaving humans across the lifespan. In recent years, fMRI studies of individual differences in clinically meaningful domains have proliferated, alongside expectations for clinical applications [1]. This expansion of individual-differences research using fMRI has triggered questions about its readiness to fulfill the measurement properties necessary for **clinical translation** (see **Glossary**), central among which is **reliability**.

**Psychometrics** has long established that reliability is the necessary first step toward validity. For example, to investigate how brain function makes a super-ager resilient in the face of neurodegeneration or to tailor brain stimulation to an individual's unique **functional topography**, we must first be able to reliably **measure** idiosyncrasies in brain function. To establish reliability, repeated measurements of brain function must produce converging estimates in the absence of significant changes in the individual (e.g., disease progression, exposure to treatment). Recently, we reported that many of the most widely adopted task-fMRI measures of brain activity during clinically relevant behavior (e.g., episodic memory, executive control) have low test-retest reliability and are, therefore, unable to serve as clinical biomarkers in their current state [2]. Results from similar studies have also pointed to low reliability in other widely used fMRI measures including functional connectivity measures generated from short scans [3–5]. Fortunately, methods to improve reliability have long been developed and employed in the allied field of psychology (e.g., personality or cognitive assessments). However, these methods have yet to be fully adopted in fMRI research.

In this review, we begin by describing historical trends that contributed to the widespread use of unreliable fMRI measures in individual-differences research. Then we highlight four

\*Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

\*Correspondence:  
maxwell.elliott@duke.edu (M.L. Elliott).

# Measurement Reliability: Big Data Resource

An open science resource for establishing reliability and reproducibility in functional connectomics

Xi-Nian Zuo<sup>1,2,\*</sup>, Jeffrey S. Anderson<sup>3</sup>, Pierre Bellec<sup>4</sup>, Rasmus M. Birn<sup>5</sup>, Bharat B. Biswal<sup>6</sup>, Janusch Blautzik<sup>7</sup>, John C.S. Breitner<sup>8</sup>, Randy L. Buckner<sup>9</sup>, Vince D. Calhoun<sup>10</sup>, F. Xavier Castellanos<sup>11,12</sup>, Antao Chen<sup>13</sup>, Bing Chen<sup>13</sup>, Jiangtao Chen<sup>1</sup>, Xu Chen<sup>1</sup>, Stanley J. Colcombe<sup>14</sup>, William Courtney<sup>15</sup>, R. Cameron Craddock<sup>13,14</sup>, Adriana Di Martino<sup>12</sup>, Hao-Ming Dong<sup>1,15</sup>, Xiaolan Fu<sup>1,16</sup>, Qiyong Gong<sup>17</sup>, Krzysztof J. Gorgolewski<sup>18</sup>, Ying Han<sup>19</sup>, Ye He<sup>1,13,15</sup>, Yong He<sup>19</sup>, Erica Ho<sup>1,14,15</sup>, Avram Holmes<sup>23</sup>, Xiao-Hui Hou<sup>1,15</sup>, Jeremy Huckins<sup>19</sup>, Tianzi Jiang<sup>23</sup>, Yi Jiang<sup>9</sup>, William Kelley<sup>23</sup>, Clare Kelly<sup>13</sup>, Margaret King<sup>19</sup>, Stephen M. LaConte<sup>1</sup>, Janet E. Lainhart<sup>1</sup>, Xu Lei<sup>1</sup>, Hui-Jia Li<sup>1</sup>, Kaiming Li<sup>1</sup>, Kuncheng Luna<sup>1</sup>, Qixiang Lin<sup>1</sup>, Dongqiang Liu<sup>1,3</sup>, Jia Liu<sup>1</sup>, Xun Liu<sup>1</sup>, Yijun Liu<sup>1</sup>, Guangming Lu<sup>16</sup>, Jie Lu<sup>2</sup>, Beatriz Luna<sup>1</sup>, Jim Luo<sup>15</sup>, Daniel Lurie<sup>1,14</sup>, Ying Mao<sup>1</sup>, Daniel S. Margulies<sup>15</sup>, Andrew R. Mayer<sup>19</sup>, Thomas Meindl<sup>1</sup>, Mary E. Meyerand<sup>1</sup>, Weizhi Nan<sup>1,25</sup>, Jared A. Nielsen<sup>1</sup>, David O'Connor<sup>1,24</sup>, David Paulsen<sup>1</sup>, Vivek Prabhakaran<sup>15</sup>, Zhigang Qi<sup>20</sup>, Jiang Guo<sup>1</sup>, Chunhong Shao<sup>15</sup>, Zarar Shehzad<sup>13,14</sup>, WeiJian Tang<sup>1</sup>, Arno Villringer<sup>23</sup>, HuaLing Wang<sup>1</sup>, Kai Wang<sup>1,15</sup>, Dongtao Wei<sup>1</sup>, Gao-Xia Wei<sup>1</sup>, Xu-Chu Weng<sup>1</sup>, Xuehai Weng<sup>19</sup>, Ting Xu<sup>1,13,15</sup>, Ning Yang<sup>1,15</sup>, Zhi Yang<sup>1</sup>, Yu-Feng Zang<sup>13</sup>, Lei Zhang<sup>1,15</sup>, Qinglin Zhang<sup>1</sup>, Zhe Zhang<sup>1,15</sup>, Zhiqiang Zhang<sup>16</sup>, Ke Zhao<sup>1</sup>, Zonglei Zhen<sup>19</sup>, Yuan Zhou<sup>1</sup>, Xing-Ting Zhu<sup>1,15</sup> & Michael P. Milham<sup>13,14</sup>



**ADVANCES IN NEURODEGENERATIVE AND PSYCHIATRIC IMAGING SPECIAL FEATURE: REVIEW ARTICLE**  
**Neuroimaging in Psychiatry and Neurodevelopment: why the emperor has no clothes**

<sup>1</sup>ASHLEY N. ANDERSON, <sup>2</sup>JACE B. KING, PhD and <sup>2</sup>JEFFREY S ANDERSON, MD PhD



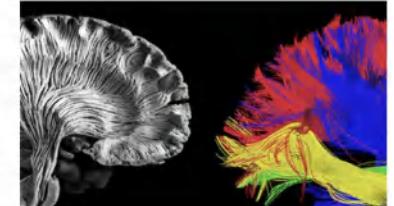
国际信度与可重复性联盟-18个中心，共享1629人的逾万套重复MRI测量

*Scientific Data* (2014); *British Journal of Radiology* (2019)

COLLECTION | 20 JANUARY 2015

## Human brain MRI reproducibility

This collection presents a series of articles describing human brain scans – produced with a variety of magnetic resonance imaging (MRI) methods and modalities – which are designed to help researchers assess the reproducibility of brain imaging techniques and to develop new methods based on these data-types. Central to this collection are studies from the Consortium on Reliability and Reproducibility (CoRR), an initiative that has organized the release of data from thousands of individual brain scans collected at 18 international sites. show less



## Author's corner: A testbed for reproducible and standardized human MRI connectomics

September 6, 2016 | 2:00 pm | Posted by Varsha Khodiyar | Category: Author's Corner, Guest Posts

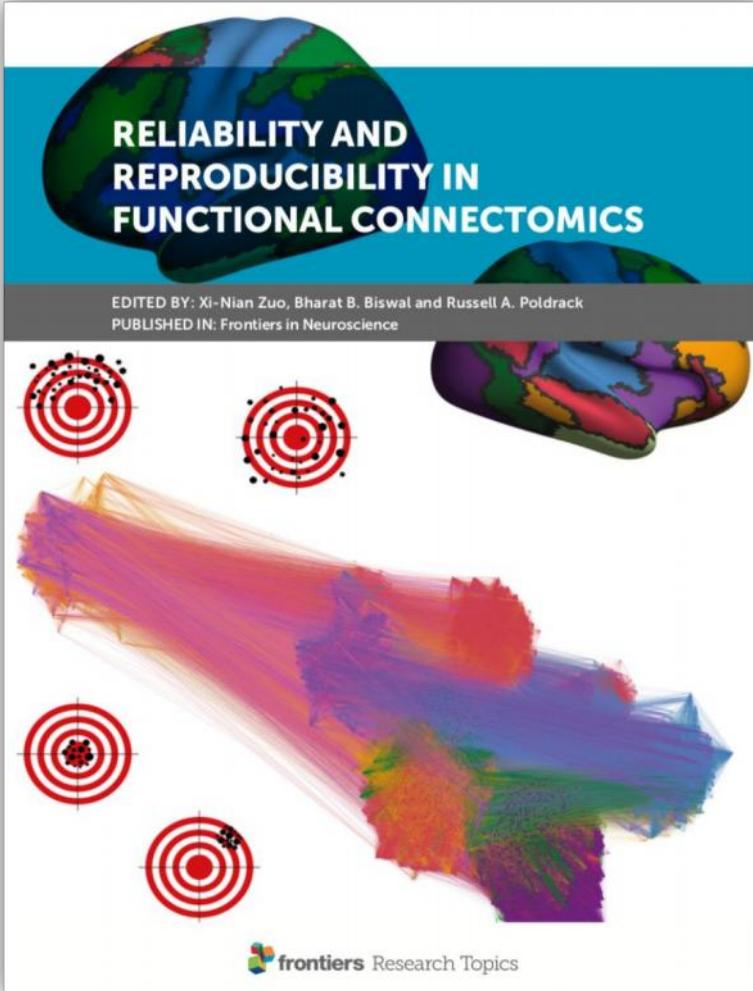
Guest post by Xi-Nian Zuo, Project Coordinator and Co-Founder of Consortium for Reliability and Reproducibility (CoRR), Professor of Psychology and Director of the Magnetic Resonance Imaging Research Center in the Institute of Psychology at Chinese Academy of Sciences, China.

About a decade ago (2006), as a PhD student graduating from the School of Mathematics at Beijing Normal University, I stepped into the field of neuroimaging of the human brain by way of a short job interview offered by Dr. Yu-Feng Zang, my postdoc mentor in China. The most important thing that I learned and developed during my post doc training was how to question a study, an indication likely of my somewhat different background (mathematics versus brain sciences). Probability and statistics became my major tools in bridging new learning experiences with my existing knowledge, pushing me to further pursue research training offered by Dr. Michael Peter Milham at New York University. Ongoing work in his laboratory really interested me, particularly test-retest reliability of resting-state functional connectivity<sup>1</sup>, the first study of test-retest reliability in the nascent field of functional connectivity. However, an obvious limitation existed to that study, and a series of test-retest reliability studies I carried out subsequently<sup>2</sup>, the small sample size. This directly motivated me to seek and build up a truly big data set for test-retest reliability in connectomics.



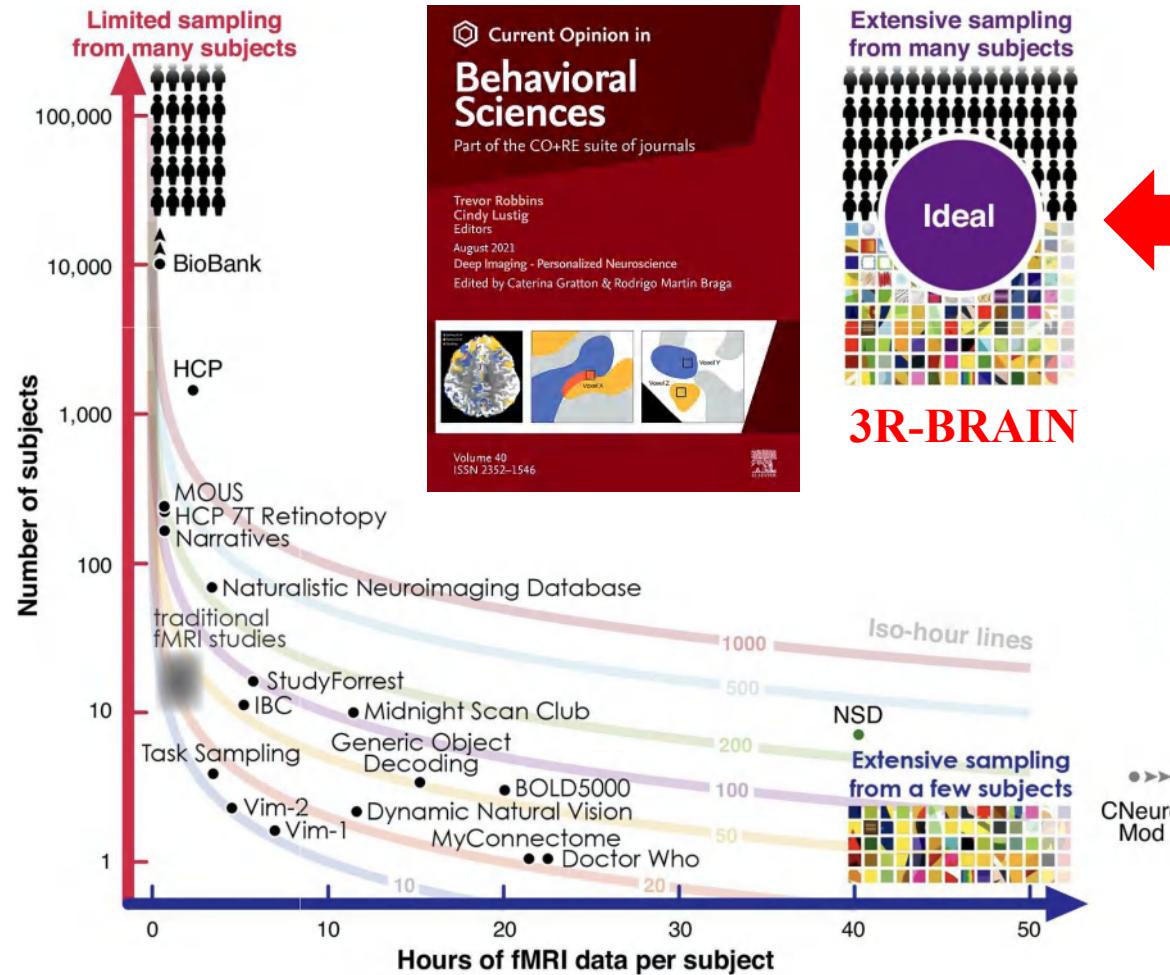
XI-NIAN ZUO

# Measurement Reliability: Big Data Resource

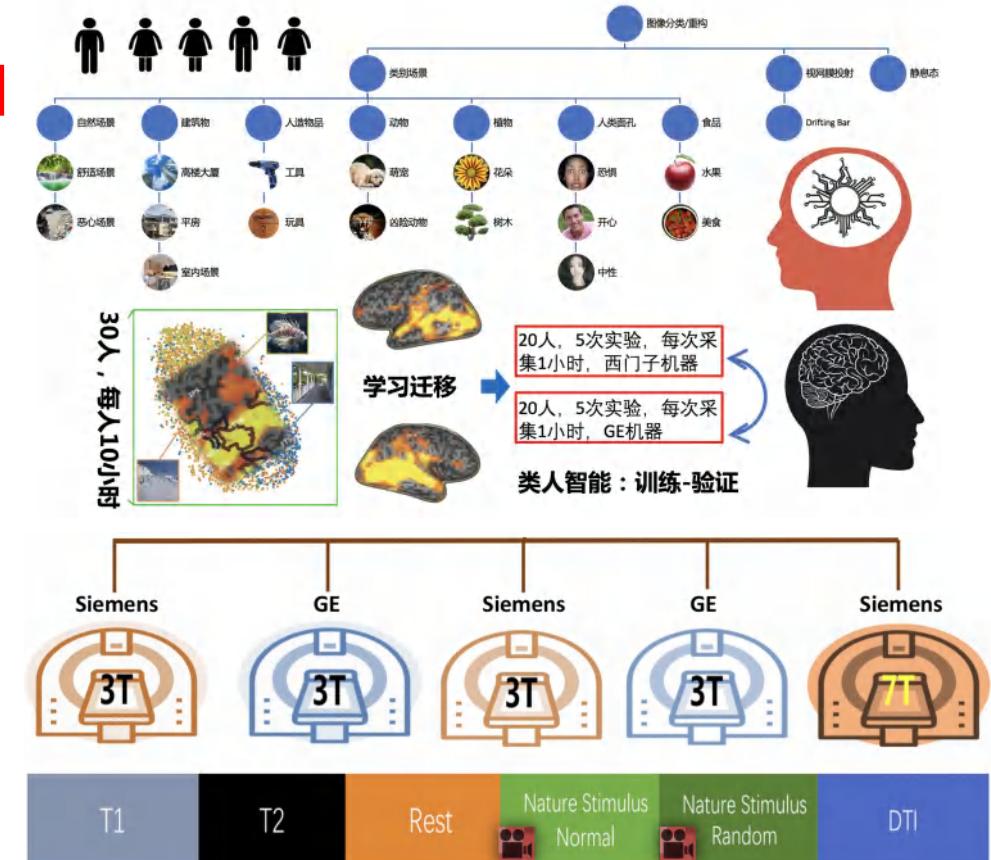


The image is a screenshot of a website page for a research topic. The header is purple with white text. It reads "Reliability and Reproducibility in Functional Connectomics" and "Research Topic". Below the header is a sub-header "Reliability and Reproducibility in Functional Connectomics". The URL "https://www.frontiersin.org/research-topics/5137/reliability-and-reproducibility-in-functional-connectomics" is displayed. To the right, there is a dark blue sidebar with "87k Views". Below the header, there are tabs for "Overview", "Articles 14", "Authors 58", and "Impact". The main content area has a heading "About this Research Topic" and a "Show more" button. At the bottom, there is a section titled "Topic Editors" featuring three profiles: Xi-Nian Zuo (Beijing Normal University), Bharat B Biswal (Department of Biomedical Engineering, Newark College of...), and Russell A Poldrack (Stanford University).

# Measurement Theory: Seeking Big Data Resource



A Brain Consortium for Reproducibility, Reliability and Replicability



# Measurement Theory Meets Big Data Science



## A hitchhiker's guide to working with large, open-source neuroimaging datasets

Corey Horien<sup>1,2,✉</sup>, Stephanie Noble<sup>3</sup>, Abigail S. Greene<sup>1,2</sup>, Kangjoo Lee<sup>2,3</sup>, Daniel S. Barron<sup>4</sup>, Siyan Gao<sup>5</sup>, David O'Connor<sup>6</sup>, Mehravesh Salehi<sup>5,6</sup>, Javid Dadashtkarimi<sup>7</sup>, Xilin Shen<sup>3</sup>, Siyuan M. R. Lake<sup>3</sup>, R. Todd Constable<sup>8,9,10</sup> and Dustin Scheinost<sup>1,3,5,6,✉</sup>

Large datasets that enable researchers to perform investigations with unprecedented rigor are growing increasingly common in neuroimaging. Due to the simultaneous increasing popularity of open science, these state-of-the-art datasets are more accessible than ever to researchers around the world. While analysis of these samples has pushed the field forward, they pose a new set of challenges that might cause difficulties for novice users. Here we offer practical tips for working with large datasets from the end-user's perspective. We cover all aspects of the data lifecycle: from what to consider when downloading and storing the data to tips on how to become acquainted with a dataset one did not collect and what to share when communicating results. This manuscript serves as a practical guide one can use when working with large neuroimaging datasets, thus dissolving barriers to scientific discovery.

As a part of the open science movement in neuroimaging, many large-scale datasets, including the Human Connectome Project (HCP), the Adolescent Brain Cognitive Development (ABC) study<sup>1</sup> and the UK Biobank<sup>2</sup>, have been released to investigators around the world (Fig. 1; abbreviations for datasets provided in Supplementary Table 1)<sup>3–5</sup>. These initiatives build upon efforts dating back to the early twentieth century to collate large-scale brain datasets (for example, ref. <sup>6</sup>) and have advanced efforts to understand human brain function. Notably, they have been realized in response to—and helped provide support for—the realization that many questions in the field are associated with small effect sizes only detectable with large samples<sup>7,8</sup>. Since adequately large samples can be difficult for any single lab to collect in isolation, these large datasets unlock a path to investigate previously inscrutable questions.

Nevertheless, use of these large datasets can be daunting. With thousands of participants and substantial imaging data per individual, simply downloading and storing the data can be difficult. The complex structure of these large datasets (for example, multiple data releases from HCP, multiple sites contributing to ABCD, etc.) presents considerable challenges and requires adherence to best practices. Even day-to-day concerns, like maintaining a lab notebook, take on new importance when handling such data.

Here, we present tips for those who will be handling these data as end-users. We offer recommendations for the entire life cycle of data use—from downloading and storing data, to becoming acquainted with a dataset one did not collect, to reporting and sharing results (Table 1). Note that we do not provide recommendations for specific analytical approaches using large datasets, as these topics have been discussed elsewhere<sup>9–11</sup>. Our intention is to bring together in one place accessible and general recommendations, incorporating

practical suggestions based on our experience working with numerous large datasets. Our intended reader is one who might be associated with working with a large dataset for the first time, and we envision this manuscript to serve as an ongoing guide throughout this exciting process.

### Obtaining and managing data

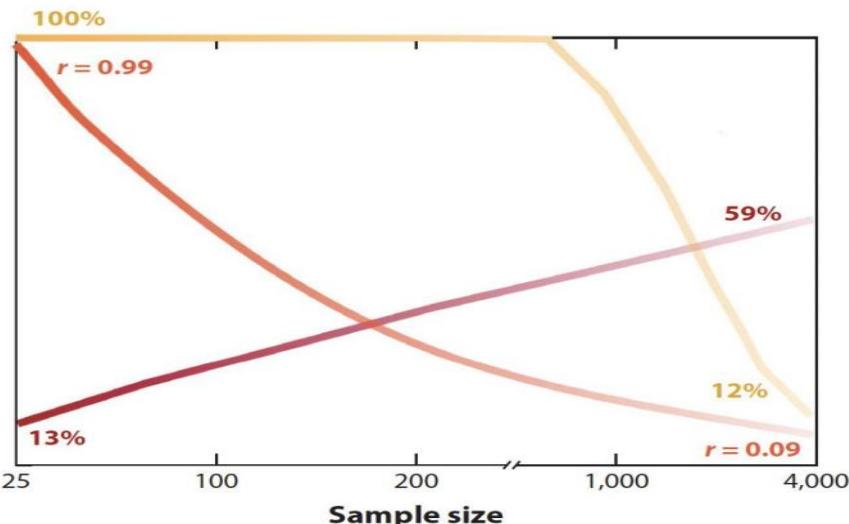
In the first section, we discuss obtaining and managing large datasets. Careful planning can help ensure that preprocessing and analysis goes smoothly, saving time in the future.

**Identifying research questions.** Given that large, open-source datasets consist of many different types of data, the first step is identifying the dataset that can address a study's question of interest. Most large datasets have some combination of imaging, genetic, behavioral and other phenotypic data (Fig. 1) that may not be harmonized across different datasets. Some may include specific clinical populations and/or related measures. To more robustly address the research question, a researcher may leverage multiple datasets to bolster sample sizes (i.e., for a rare subset of the data or for participants with a rare disease) or to demonstrate reproducibility of findings across samples. Whatever the intended use, giving careful thought to the scientific question at hand will help focus the researcher and identify which types of data are needed. At this stage, investigators may also wish to preregister their research question and analysis plans. In addition, it is important to consider the original purpose of the dataset, as it might influence the sort of questions that can be addressed, as well as the feasibility of using it in conjunction with other open-source datasets. Indeed, understanding the original purpose of a dataset can facilitate analyses, including in some cases analyses performed many years after the original data were collected<sup>12</sup>.



*Annual Review of Developmental Psychology*

## Developmental Cognitive Neuroscience in the Era of Networks and Big Data: Strengths, Weaknesses, Opportunities, and Threats



<sup>1</sup>Interdepartmental Neuroscience Program, Yale School of Medicine, New Haven, CT, USA. <sup>2</sup>MD/PhD program, Yale School of Medicine, New Haven, CT, USA. <sup>3</sup>Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA. <sup>4</sup>Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. <sup>5</sup>Department of Biomedical Engineering, Yale University, New Haven, CT, USA. <sup>6</sup>Summary Analytics Inc., Seattle, WA, USA. <sup>7</sup>Department of Computer Science, Yale University, New Haven, CT, USA. <sup>8</sup>Department of Neurosurgery, Yale School of Medicine, New Haven, CT, USA. <sup>9</sup>Department of Statistics & Data Science, Yale University, New Haven, CT, USA. <sup>10</sup>Child Study Center, Yale School of Medicine, New Haven, CT, USA. <sup>✉</sup>E-mail: corey.horien@yale.edu; dstein@yale.edu

# Measurement Theory Meets Big Data Science

**ANNUAL REVIEWS**

*Annual Review of Developmental Psychology*  
Developmental Cognitive Neuroscience in the Era of Networks and Big Data: Strengths, Weaknesses, Opportunities, and Threats

Damien A. Fair,<sup>1,2,3</sup> Nico U.F. Dosenbach,<sup>4</sup> Amy H. Moore,<sup>3,5</sup> Theodore D. Satterthwaite,<sup>6</sup> and Michael P. Milham<sup>7,8</sup>

<sup>1</sup>Institute of Child Development, College of Education and Human Development, University of Minnesota, Minneapolis, Minnesota 55455, USA; email: fair@umn.edu  
<sup>2</sup>Department of Pediatrics, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA  
<sup>3</sup>Masonic Institute for the Developing Brain, University of Minnesota, Minneapolis, Minnesota 55414, USA  
<sup>4</sup>Department of Neurology, Washington University School of Medicine, St. Louis, Missouri 63110, USA; email: dosenbach@wustl.edu  
<sup>5</sup>Carson School of Management Medical Industry Leadership Institute, University of Minnesota, Minneapolis, Minnesota 55414, USA; email: moore999@umn.edu  
<sup>6</sup>Penn Lifespan Informatics and Neuroimaging Center and Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: satterth@pennmedicine.upenn.edu  
<sup>7</sup>Center for Biomedical Imaging and Neuromodulation, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, New York 10962, USA  
<sup>8</sup>Child Mind Institute, New York, NY 10022, USA; email: michael.milham@childmind.org

Annu. Rev. Dev. Psychol., 2021. 3:249–75. Downloaded from www.annualreviews.org.  
Access provided by Beijing Normal University on 09/2022. For personal use only.

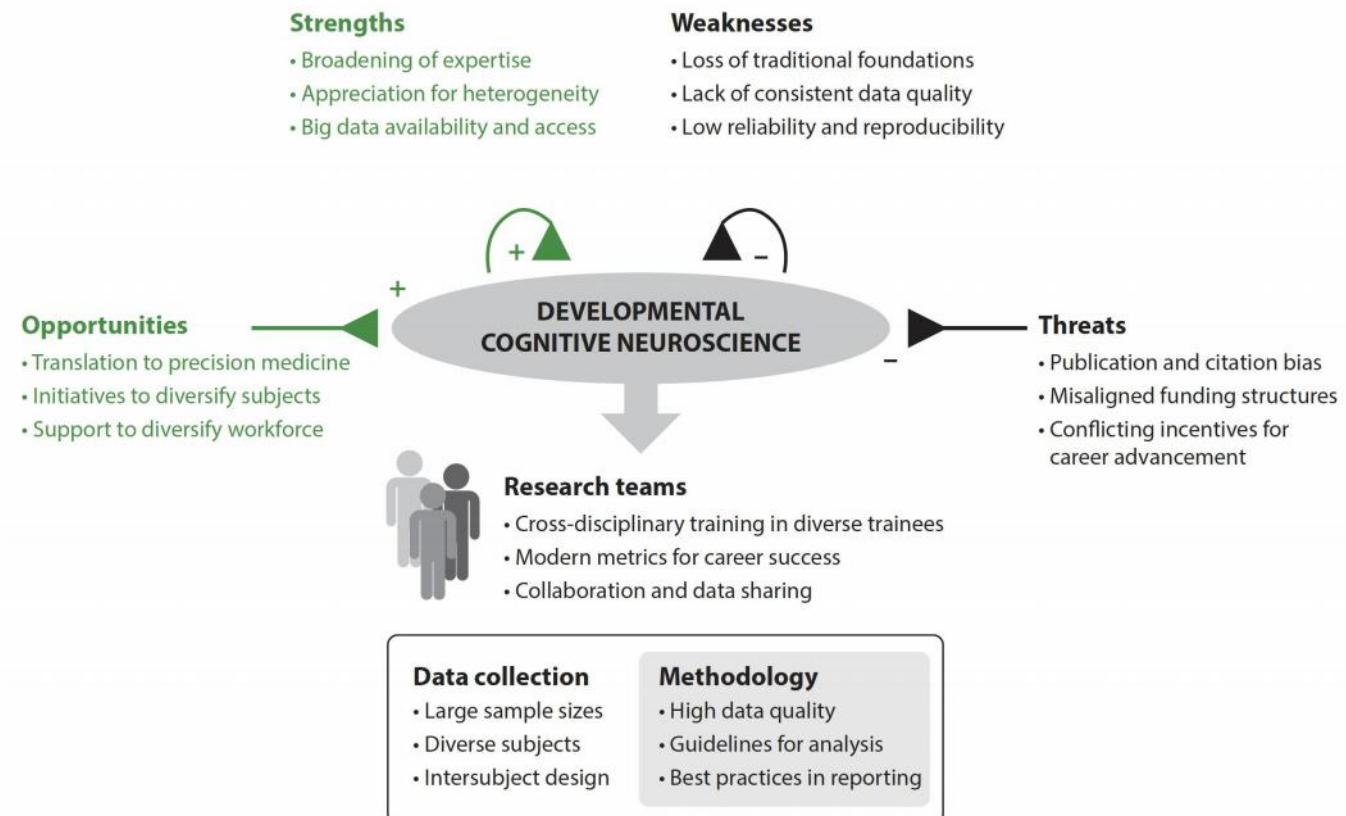
**ANNUAL REVIEWS CONNECT**  
[www.annualreviews.org](http://www.annualreviews.org)  
• Download figures  
• Navigate cited references  
• Keyword search  
• Explore related articles  
• Share via email or social media

Annu. Rev. Dev. Psychol., 2021. 3:249–75  
First published as a Review in Advance on September 30, 2021  
The *Annual Review of Developmental Psychology* is online at [depsych.annualreviews.org](https://depsych.annualreviews.org)  
<https://doi.org/10.1146/annurev-devpsych-121318-085124>  
Copyright © 2021 by Annual Reviews.  
All rights reserved

**Keywords**  
network neuroscience; development; functional connectivity; brain; cognition; strengths, weaknesses, opportunities, and threats; SWOT

**Abstract**  
Developmental cognitive neuroscience is being pulled in new directions by network science and big data. Brain imaging [e.g., functional magnetic resonance imaging (fMRI), functional connectivity MRI], analytical advances (e.g., graph theory, machine learning), and access to large computing resources have empowered us to collect and process neurobehavioral data

249



# Measurement Theory Meets Big Data Science

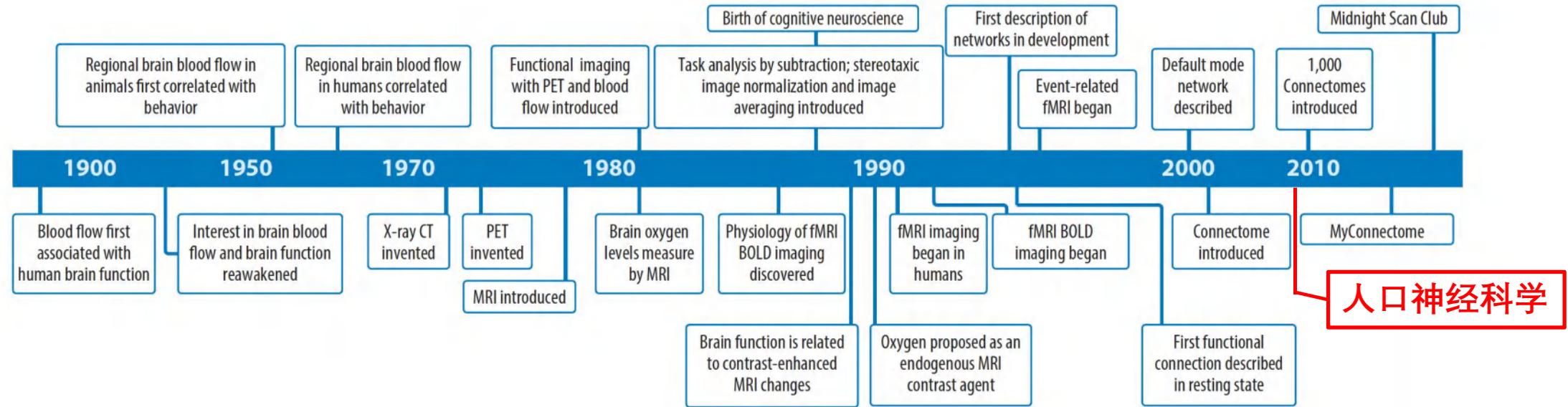


Figure 1

Timeline of developments in brain mapping and network neuroscience. Abbreviations: BOLD, blood oxygenation level-dependent; CT, computerized tomography; fMRI, functional magnetic resonance imaging; MRI, magnetic resonance imaging; PET, positron emission topography. Figure adapted with permission from Raichle (2009).

发展人口神经科学：网络与大数据时代的发展认知神经科学