The Statistical Analysis of the Varying Brain

Oliver Y. Chén, Duy Thanh Vũ, Gilbert Greub, Hengyi Cao, Xingru He, Yannick Muller, Constantinos Petrovas, Haochang Shou, Viet-Dung Nguyen, Bangdong Zhi, Laurent Perez, Jean-Louis Raisaro, Guy Nagels, Maarten de Vos, Wei He, Raphael Gottardo, Palie Smart, Marcus Munafò, Giuseppe Pantaleo

***Abstract*—We present here a systematical approach to studying the varying brain. We first distinguish different types of brain variability and provide examples for them. Next, we show classical analysis of covariance (ANCOVA) as well as advanced residual analysis via statistical- and deep-learning aim to decompose the total variance of the brain or behaviour data into explainable variance components. Additionally, we discuss innate and acquired brain variability. For varying *big brain data*, we define the *neural law of large numbers* and discuss methods for extracting representations from large-scale, potentially high-dimensional brain data. Finally, we examine the gut-brain axis, an often lurking, yet important, source of brain variability.**

***Index Terms*—Brain variability, innate variability, acquired variability, *Bayesian* brain, ANCOVA, residual learning, high-dimensional data, gut-brain axis.**

I. Prologue

In *On the Origin of Species*, Darwin discussed the importance of variability and argued it is greatest in structures that evolve fastest [1]. In humans, the brain is the most variable organ [2]. The inspection of cytoarchitecture by Campbell and Brodmann unveiled the brain's varying organization and functioning [3], [4]. The concept of *variance* introduced by Fisher facilitated the quantitative enquiry of biological variability [5]. In the past century, the study of brain variability has uncovered fresh insights about the brain, mind, and behaviour [6]–[11]. Meanwhile, linking (co)varying neural features with cognitive, behavioural, and disease outputs has revealed plausible neural origins of cognition and attention [12], [13] and potential markers predictive of brain disorders [14], [15].

This piece aims to connect two areas of cardinal, equal importance: (1) the biological variability of the brain and (2) the statistical methods useful to study it. To do so, we present the following topics. (i) We define different types of brain variability. (ii) We explore how to decompose the brain's total variance into sensible variance components. (iii) We argue a distinction between innate and acquired brain variability. (iv) We suggest methods to obtain representations of varying high-dimensional brain data. (v) We discuss brain variability due to the gut-brain axis.

OY Chén is with Centre Hospitalier Universitaire Vaudois (CHUV) and Université de Lausanne (olivery.chen@chuv.ch). H Cao is with Northwell Health. G Greub, Y Muller, C Petrovas, L Perez, JL Raisaro, R Gottardo, and G Pantaleo are with the CHUV. X He and W He are with the He University and He Eye Specialist Hospital, respectively. DT Vũ and VD Nguyen are with AVITECH. H Shou is with the University of Pennsylvania. B Zhi, P Smart, and M Munafò are with the University of Bristol. G Nagels is with Oxford and VUB. M de Vos is with KU Leuven. DT Vũ processed data for Fig 1. H Cao processed data for Fig 2. G Greub wrote Section VI and conceptualised Fig 5. The authors thank H Phan for comments on Sec. III C.

We hope our explorations may stir further discussions about neurobiological underpinnings of brain variability and help develop reliable and reproducible methods to study it.

II. Defining brain variability

A powerful way to decipher the varying brain is to see it as a spatial matrix travelling in time (**Figure** 1 a). Let $y_i(v,t)$ be the brain data[1] measured at area $v \in \{1,\dots,V\}$ at time $1 \le t \le T$ from an individual $1 \le i \le N$. Following subject i's area v over time, the trajectory $\{y_i(v,1),\dots,y_i(v,T)\}$ shows temporal variation (top panel of **Figure** 1 b). Fixing time t, the distribution $\{y_i(1,t),\dots,y_i(V,t)\}$ across areas depicts spatial variation (each single brain in the top panel of **Figure** 1 b). Fixing space v and time t, the distribution $\{y_1(v,t),\dots,y_N(v,t)\}$ presents within-group variation (*e.g.*, HC, MCI, and AD groups in **Figure** 1 b). Finally, patterns between groups (*e.g.*, male $\{y_1^M(v,t),\dots,y_{N_1}^M(v,t)\}$ *vs.* female $\{y_1^F(v,t),\dots,y_{N_2}^F(v,t)\}$, or healthy *vs.* disease) mark the between-group variation (**Figure** 1 c-d).

Differentiating brain variability launches specialized brain studies. Temporal variability: Tracking temporal brain variability, one gains insights into time-varying neural dynamics (*e.g.*, the "dynamic core" [18]), neural development, and brain maturation [16], [19], [20]. Analysing temporal brain variability for elderlies or patients with neurodegenerative diseases helps understand the ageing brain [21] and the ailing brain [22]. Additionally, past neural activities help make forecasts about future activities [23]. Spatial variability: Studying brain signals co-vary in space helps decrypt how the brain is wired [24], [25]. Within- and between-group variability: Examining brain patterns within and across groups, one derives population-level characteristics [26] and subject-specific information [14]. Potential causal variability: Examining brain patterns co-varying with stimuli and/or behaviour helps identify neural signatures: for processing the stimuli [27], predictive of cognition [28], and intermediating stimuli and behaviour [29].

III. Identifying and quantifying brain variability

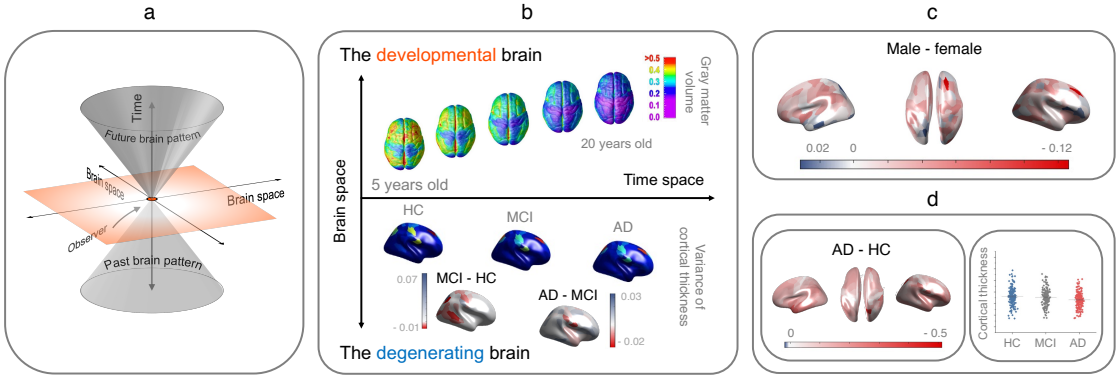
One way to trace what constitutes brain variability is to decompose its total variability into parts related to brain-space, time, individual, age, gender, *etc*. Here, we present a few ways to identify, isolate, and quantify variance components.

A. Variance decomposition via ANCOVA

Consider subject i ($1 \le i \le N = 165$; 51 HCs and 114 patients with psychosis)'s brain region j ($1 \le j \le J = 13$). Let

[1]We consider the location and types of signals broadly: the former can be a neuron, a voxel, or a brain parcel; the latter can be the action potentials of single neurons, BOLD fMRI of voxels, or EEG recordings of electrodes.

978-1-6654-5245-8/23/$31.00 ©2023 IEEE



(a) **The spatial-temporal brain.** An observer is at the centre of a brain matrix at $t = 0$. Looking around, the observer sees brain activities across the brain at present. Looking below and above, one perceives past and future brain activities. (b) The **developmental** and the **degenerating** brain. (c) Cortical thickness between **males** and **females**. (d) Cortical thickness between **disease groups**. Data for top panel of (b) are from [16]. AD data are from the *Alzheimer's Disease Neuroimaging Initiative* (ADNI) [17]. HC = healthy control ($N = 334$); MCI = mild cognitive impairment ($N = 591$); AD = Alzheimer's ($N = 182$).

Fig. 1. Distinguishing types of brain variability.

k ($200 \leq k \leq 428$), $l \in \{\text{HC}, \text{Psychosis}\}$, and $g \in \{M, F\}$ be the subject's age in months, disease status, and gender, respectively. Each subject is measured at time (visit) t ($1 \leq t \leq T_i = 2$). Then, one can write the brain data (*e.g.*, mean functional connectivity of region j) from subject i (with age k, disease status ℓ, and gender g) measured at time t, or $y^{(j)}_{iklg}(t)$, as:

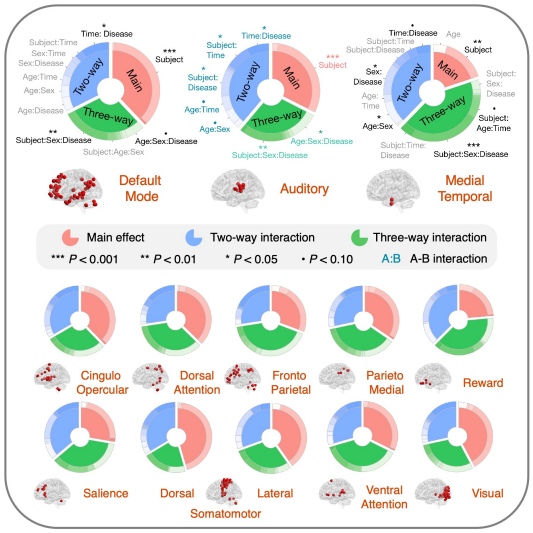
$$y^{(j)}_{iklg}(t) = \mu^{(j)} + \sum_{\substack{w\in \\ \{i,t,k,\ell,g\}}} \beta^{\{j\}}_{w} + \sum_{\substack{u\neq v \\ u,v\in \\ \{i,t,k,\ell,g\}}} \alpha^{(j)}_{u,v} + \sum_{\substack{a\neq b, a\neq c, b\neq c \\ a,b,c\in \\ \{i,t,k,\ell,g\}}} \gamma^{(j)}_{a,b,c} + \epsilon^{(j)}_{iklg}(t)$$

where, for area j, $\mu^{(j)}$ is the population mean, $\beta^{(j)}_i$, $\beta^{(j)}_t$ $\beta^{(j)}_k$, $\beta^{(j)}_l$, $\beta^{(j)}_g$ are the main-effects due to individual i, time t, age k, disease l, and gender g; $\alpha^{(j)}_{u,v}$'s are two-way interactions; $\gamma^{(j)}_{a,b,c}$'s are three-way interactions (interactions beyond three-way is usually very small [30]); $\epsilon^{(j)}_{iklg}(t)$ is the residual.

The sum of squared terms of the ANCOVA can then be used to analyze how the main effects and interactions may contribution to the total variance (see **Figure** 2).

B. Further variance decomposition on the residuals

The residuals $\left\{\epsilon^{(j)}_{iklg}(t)\right\}$ still contain unexplained information. Extending the ANCOVA model to a general form, one can further study the residual. Formally, $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the residual following $MVN(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \text{blockdiag}\{V_1, \ldots, V_N\}$, and V_i models the within-subject correlation, for $1 \leq i \leq N$. The model is relatively general: if $\boldsymbol{Y}$ is the brain data, $\boldsymbol{X}$ can incorporate individual, temporal, spatial, interaction, and covariates; if $\boldsymbol{Y}$ is measured behaviour, and one wants to study how brain signals affect $\boldsymbol{Y}$ while controlling for other effects, $\boldsymbol{X}$ can include both brain signals and other factors. One then decomposes $\boldsymbol{\epsilon}$ into random effect,



Variance decomposition on functional brain connectivities. The proportion of residual variance is not shown. Data are from the *Human Connectome Project for Early Psychosis* (51 HCs and 114 patients with psychosis; Males/Female/*NA*: 101/62/2) [31].

Fig. 2. Decomposing the variance of the brain.

serial correlation, and measurement error (**Table** 2 in [32]), and interprets the residual variability using further components.

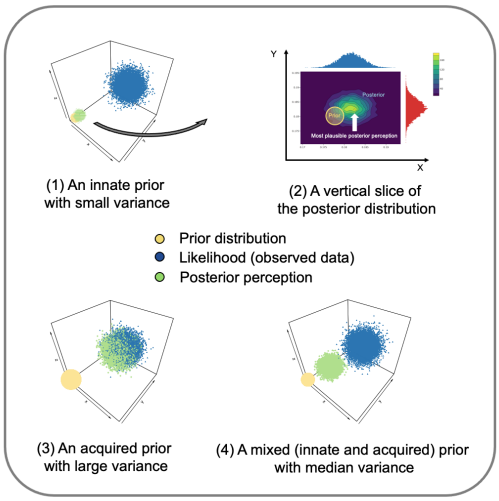
C. Variance decomposition on the residuals via deep learning

If $\boldsymbol{X}$ relates to $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$ in a non-linear, potentially multi-layer way, one can first employ a deep neural network to map $\boldsymbol{X}$ into a feature representation space, where $\boldsymbol{X}$'s representations are entangled in two subspaces: one with information explainable for $\boldsymbol{Y}$ and the other explainable for $\boldsymbol{\epsilon}$; one then separates the two subspaces. For the first subspace, one designs a network to predict $\boldsymbol{Y}$ as accurately as possible, while forcing the prediction accuracy of $\boldsymbol{\epsilon}$ to chance level, for example, via a gradient reversal layer [33]. For the second subspace, one

devises the network to predict $\boldsymbol{\epsilon}$ as accurately as possible while forcing the prediction accuracy of $\boldsymbol{Y}$ to chance level. Formally, the network learns to minimize the loss function:

$$\mathscr{L} = [\mathscr{L}(\boldsymbol{Y}|F_{\boldsymbol{Y}}) - \mathscr{L}(\boldsymbol{\epsilon}|F_{\boldsymbol{Y}})] + [\mathscr{L}(\boldsymbol{\epsilon}|F_{\boldsymbol{\epsilon}}) - \mathscr{L}(\boldsymbol{Y}|F_{\boldsymbol{\epsilon}})]$$

where $F_{\boldsymbol{Y}}$ and $F_{\boldsymbol{\epsilon}}$ denote the feature subspaces solely explaining for $\boldsymbol{Y}$ and $\boldsymbol{\epsilon}$, respectively.



(1) The **innate prior** with small variability (*i.e.*, high precision). The posterior learns mainly from the prior; not much from the observations. (2) With an innate prior, the centre of the posterior is close to that of the prior. (3) The **acquired prior** with large variability. The posterior learns mainly from the observations. (4) When a prior has moderate variability, the posterior learns from the prior *and* the observations. Adapted from Zeki and Chén [34].

Fig. 3. Two tales of brain variability.

IV. Innate and Acquired Brain Variability

The brain's structure and functioning are, in part, decreed by innate factors, and, in part, adaptive to the external world [19], [34]. But why and how do we distinguish innate and acquired brain variability? We draw insights from three directions.

From biology to neurobiology. One needs to distinguish innate and acquired brain variability [34]. The former is likely dictated by the genes and less variable (compared to the latter). The latter is developed postnatally, due to environment or a combination of environmental and genetic factors and is more variable. Next is to find evidence. Colour perception varies little in humans: perceiving white colour after seeing a white flag is independent of culture and learning [35]. Perceiving ceasefire when seeing a white flag, however, depends on postnatal learning, and is variable across cultures [34].

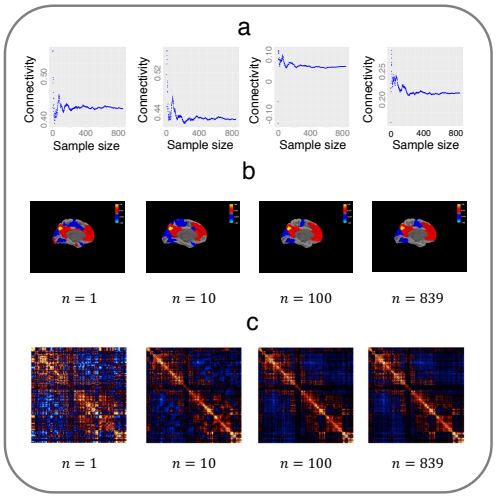
From statistics to neurobiology. Variance decomposition (**Section** III) separates the total variability of the (neural and behavioural) phenotypes into genetic, environmental, and (gene-gene, environment-environment [36], and gene-environment [37]) interaction parts. A derivative of variance decomposition is the heritability ($H^2 := V(G)/V(P)$), namely the genetic variance over the total variance.

From innate and acquired brain variability to the *Bayesian* brain. Empirical studies suggest the brains of small animals [38], mammals [39], and monkeys [40] perform *Bayesian* integration and updating. Humans, too, seem to use probability updating to modify perception [41], cognition [42], and sensorimotor function [43]. Incorporating prior knowledge (encoded in genes and acquired through learning) with new information, one updates perception and behaviour with higher precision (*i.e.*, lower variation) [34] (see **Figure** 3).

V. Analysis of Variability Using Big Brain Data

A. The neural law of large numbers (NLLN)

Empirical analyses of brain connectivity hint the ***neural law of large numbers*** (NLLN) (**Figure** 4). Formally, we define the NLLN as the principle where the averaged brain signals converges *asymptotically* as the sample size increases. The law also implies that group-level estimates from small samples may fluctuate; as the sample size increases, the estimates become stable and may offer improved population-level insights and perhaps general principles of the brain.



Top to bottom are connectivities between paired brain areas, between a seed (in the posterior cingulate; yellow coulour) and the rest of the brain, and the whole-brain (400 areas). The results **stabilize asymptotically**. Data are from the *Human Connectome Project* [44].

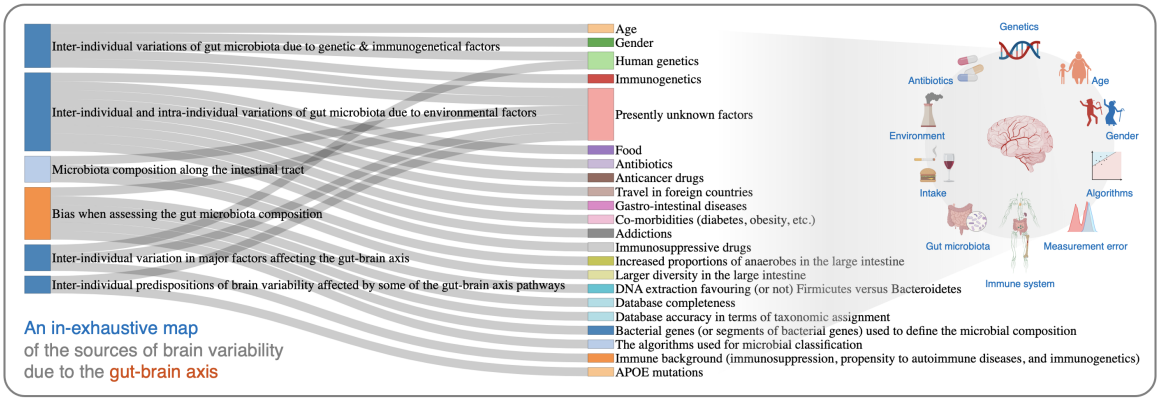
Fig. 4. The Neural Law of Large Numbers (NLLN).

B. High-dimensional brain data analysis, challenges and potential solutions

High-dimensional ***big brain data*** require special treatments, particularly when the number of features (p) goes to infinity faster than, or at the same rate as, the number of sample size (N). Under these circumstances, classical statistical theories collapse. The difficulty, however, can be partially alleviated if the data have a sparse structure[2]. One can then deploy feature selection (*e.g.*, stepwise feature selection and regularization[3]

[2]Empirical studies show sparsity in cells, genes, and the brain [45].

[3]One can use *sure independence screening* when p grows faster or as fast as N, or sparse additive models and *Bayesian* additive regression trees when the relationship between brain features and the outcome is nonlinear.



Sources of variability of the brain through the **gut-brain axis** and causes of possible **biases** due to the tools used to determine the gut microbiota composition: an in-exhaustive map. Parts of the figure are drawn using sankeyNetwork and BioRender.

Fig. 5. The spell of the gut-brain axis.

[46]) or transformation (*e.g.*, eigenanalysis or seek a low-dimensional nonlinear manifold) to discover a sparse or low-dimensional representation (see [32] for discussions).

Big brain data confront further challenges. (a) Consider p brain areas. If p is large, signals from some areas may be spuriously associated with an outcome. Out-of-sample prediction, cross-validation, and repeated sampling test may alleviate this [47]. (b) Most statistical models require the predictors to be uncorrelated with the residual. When p is large, some variables may be coincidentally correlated with the residual (*i.e.*, incidental endogeneity); see [48] for a potential treatment. (c) Aggregating datasets of different noise levels may bias the estimates. One needs suitable pre-processing and aggregation methods (*e.g.*, inverse-variance weighting). (d) High-dimensional brain data analysis face computational challenges; many high-dimensional problems are intractable; they may generate even larger intermediate data. But see treatments (*e.g.*, assuming sparsity) above. (e) Data with large sample sizes may yield small, yet significant, effect sizes. P-values in these cases may offer little inference value. See [47] for discussions. (f) Data visualization is critical to exploratory analysis and *post hoc* interpretation, but plotting high-dimensional data is difficult. Instead, one can project high-dimensional data onto low-dimensional space [49].

VI. A Lurking Factor: The Gut-Brain Axis

Despite growing insights about the potential roots of brain variability due to genetics, neuro-development, cognition, stimulation, education, environmental exposure, life events, and innate immunity, one source of brain variability, namely that due to endogenous microbes, has been little explored.

Recent years, however, have seen an increasing interest in investigating the links between microbes and the variability of the brain and the inquiry of the *gut-brain axis*. Remarkable beginnings have already been made. Studies have shown that gut microbiota compositions are associated with (i) psychiatric disorders, such as depression and compulsive disorders [50], (ii) neurological diseases, such as Alzheimer-associated memory impairment, Parkinson's disease, and multiple sclerosis [51], and (iii) changes of major physiological functions, such as appetite, motivation to perform an exercise, and sleep [52].

Although the definitive causal pathways between these factors have not been fully charted, for some of these conditions, robust data have provided partial explanations about how gut microbes affect the brain. For example, gut microbes (*Eubacterium rectale* & *Coprococcus eutactus*) seem to improve the exercise performance of mice via producing endocannabinoid metabolites that indirectly impact the dopamine level in the brain [52]. Part of the impact of gut microbiota on the brain, however, was shown to depend on immune interplay with microbes. Consequently, prebiotics and probiotics as well as immunomodulators are now being investigated as possible future treatments for such psychiatric or neurological impairments.

It is as of yet elusive to precisely pinpoint the impact of gut microbiota variability on brain physiology and its intrinsic variability. This is due partly to the inter-individual variability of the gut microbiota, the intra-individual variability of gut microbiota over time and at different anatomical locations of the intestinal tract, the varied methods used to define the gut microbiota composition, and the differing effect of a given microbiota composition on the brain according to genetic predisposition and immune background. Moreover, various external factors (*e.g.*, travel, diet, antibiotics, exposure to pets, and gastrointestinal disease) may further affect and explain a portion of the variability of the gut microbiota composition and its impact on brain functions and, potentially, structure (**Figure** 5). Future work needs to identify, isolate, and quantify these entangled factors and multi-layer pathways of the gut-brain axis.

VII. Epilogue

In this piece, we have presented two views about the varying brain: a biological one and a statistical one. Future studies of variability will continue to expand our knowledge about the genetic, neural, and environmental bases of the brain, and how its consistent and varying structure and functioning interact, integrate, and shape the constant and adapting humans.

REFERENCES

[1] C. Darwin, *On the origin of species*. London, UK: John Murray, 1859.

[2] S. Zeki, "Artistic creativity and the brain," *Science*, vol. 293, no. 5527, pp. 51–52, 2001.

[3] A. W. Campbell, "Histological studies on the localisation of cerebral function," *Journal of Mental Science*, vol. 50, no. 211, pp. 651–662, 1904.

[4] K. Brodman, *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig, Germany: Barth, 1909.

[5] R. A. Fisher, "The correlation between relatives on the supposition of Mendelian inheritance," *Trans. R. Soc. Edinburgh*, vol. 52, no. 2, pp. 399–433, 1918.

[6] P. Broca, "Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole)," *Bulletin et Memoires de la Societe Anatomique de Paris*, vol. 6, pp. 330–357, 1861.

[7] F. G and E. Hitzig, "Über die elektrische Erregbarkeit des Großhirns," *Physiol. und Wissenschaftliche Med*, vol. 37, pp. 300–332, 1870.

[8] C. Wernicke, *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Breslau, Poland: Cohn & Weigert, 1874.

[9] E. M. Gordon, T. O. Laumann, B. Adeyemo, and S. E. Petersen, "Individual variability of the system-level organization of the human brain," *Cerebral Cortex*, vol. 27, no. 1, pp. 386–399, 2017.

[10] M. L. Seghier and C. J. Price, "Interpreting and utilising intersubject variability in brain function," *Trends in Cognitive Sciences*, vol. 22, no. 6, pp. 517–530, 2018.

[11] S. Smith, E. Duff, A. Groves, T. E. Nichols, S. Jbabdi, L. T. Westlye, C. K. Tamnes, A. Engvig, K. B. Walhovd, A. M. Fjell *et al.*, "Structural variability in the human brain reflects fine-grained functional architecture at the population level," *Journal of Neuroscience*, vol. 39, no. 31, pp. 6136–6149, 2019.

[12] P. S. Goldman-Rakic, "Topography of cognition: parallel distributed networks in primate association cortex," *Annual Review of Neuroscience*, vol. 11, no. 1, pp. 137–156, 1988.

[13] S. De Felice and C. A. Holland, "Intra-individual variability across fluid cognition can reveal qualitatively different cognitive styles of the aging brain," *Frontiers in Psychology*, vol. 9, p. 1973, 2018.

[14] H. Cao, O. Y. Chén, Y. Chung, J. K. Forsyth, S. C. McEwen, D. G. Gee, C. E. Bearden, J. Addington, B. Goodyear, K. S. Cadenhead *et al.*, "Cerebello-thalamo-cortical hyperconnectivity as a state-independent functional neural signature for psychosis prediction and characterization," *Nature Communications*, vol. 9, no. 1, pp. 1–9, 2018.

[15] B. U. Christ, M. I. Combrinck, and K. G. Thomas, "Both reaction time and accuracy measures of intraindividual variability predict cognitive performance in Alzheimer's disease," *Frontiers in Human Neuroscience*, vol. 12, p. 124, 2018.

[16] N. Gogtay, J. N. Giedd, L. Lusk, K. M. Hayashi, D. Greenstein, A. C. Vaituzis, T. F. Nugent III, D. H. Herman, L. S. Clasen, A. W. Toga *et al.*, "Dynamic mapping of human cortical development during childhood through early adulthood," *PNAS*, vol. 101, no. 21, pp. 8174–8179, 2004.

[17] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 4, pp. 685–691, 2008.

[18] G. Tononi and G. M. Edelman, "Consciousness and complexity," *Science*, vol. 282, no. 5395, pp. 1846–1851, 1998.

[19] G. M. Edelman, *Neural Darwinism: The theory of neuronal group selection*. London, UK: Oxford University Press, 1987.

[20] B. Casey, J. N. Giedd, and K. M. Thomas, "Structural and functional brain development and its relation to cognitive development," *Biological Psychology*, vol. 54, no. 1-3, pp. 241–257, 2000.

[21] B. A. Yankner, T. Lu, and P. Loerch, "The aging brain," *Annual Review of Pathology*, vol. 3, no. 1, pp. 41–66, 2008.

[22] R. G. Burciu, E. Ofori, D. B. Archer, S. S. Wu, O. Pasternak, N. R. McFarland, M. S. Okun, and D. E. Vaillancourt, "Progression marker of Parkinson's disease: a 4-year multi-site imaging study," *Brain*, vol. 140, no. 8, pp. 2183–2192, 2017.

[23] D. Zhang, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers," *PloS One*, vol. 7, no. 3, p. e33182, 2012.

[24] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar MRI," *Magnetic Resonance in Medicine*, vol. 34, no. 4, pp. 537–541, 1995.

[25] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.

[26] M. Ingalhalikar, A. Smith, D. Parker, T. D. Satterthwaite, M. A. Elliott, K. Ruparel, H. Hakonarson, R. E. Gur, R. C. Gur, and R. Verma, "Sex differences in the structural connectome of the human brain," *PNAS*, vol. 111, no. 2, pp. 823–828, 2014.

[27] A. Rogachov, J. C. Cheng, N. Erpelding, K. S. Hemington, A. P. Crawley, and K. D. Davis, "Regional brain signal variability: a novel indicator of pain sensitivity and coping," *Pain*, vol. 157, no. 11, pp. 2483–2492, 2016.

[28] O. Y. Chén, H. Cao, J. M. Reinen, T. Qian, J. Gou, H. Phan, M. De Vos, and T. D. Cannon, "Resting-state brain information flow predicts cognitive flexibility in humans," *Scientific Reports*, vol. 9, no. 1, pp. 1–16, 2019.

[29] O. Y. Chén, C. Crainiceanu, E. L. Ogburn, B. S. Caffo, T. D. Wager, and M. A. Lindquist, "High-dimensional multivariate mediation with application to neuroimaging data," *Biostatistics*, vol. 19, no. 2, pp. 121–136, 2018.

[30] O. Y. Chén and M. Jacroux, "On the use of semi-folding in regular blocked two-level factorial designs," *Communications in Statistics-Theory and Methods*, vol. 44, no. 12, pp. 2473–2506, 2015.

[31] "Human connectome project for early psychosis," https://www.humanconnectome.org/study/human-connectome-project-for-early-psychosis, accessed: 2023-01-18.

[32] O. Y. Chén, H. Phan, G. Nagels, and M. de Vos, "On statistical analysis of brain variability," *Preprints*, p. 2020080428, 2020.

[33] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.

[34] S. Zeki and O. Y. Chén, "The Bayesian-Laplacian brain," *European Journal of Neuroscience*, vol. 51, no. 6, pp. 1441–1462, 2020.

[35] D. F. Pears, *Incompatibilities of colours*. Oxford, UK: Blackwell, 1953.

[36] C. H. Waddington, *The strategy of the genes*. Routledge, 2014.

[37] N. Valenzuela and V. Lance, *Temperature-dependent sex determination in vertebrates*. Smithsonian Books Washington, DC, 2004.

[38] W. Hunte, R. Myers, and R. Doyle, "Bayesian mating decisions in an amphipod, *Gammarus lawrencianus* Bousfield," *Animal Behaviour*, vol. 33, no. 2, pp. 366–372, 1985.

[39] T. J. Valone and J. S. Brown, "Measuring patch assessment abilities of desert granivores," *Ecology*, vol. 70, no. 6, pp. 1800–1810, 1989.

[40] H. Sohn, D. Narain, N. Meirhaeghe, and M. Jazayeri, "Bayesian computation through cortical latent dynamics," *Neuron*, vol. 103, no. 5, pp. 934–947, 2019.

[41] D. C. Knill and W. Richards, *Perception as Bayesian inference*. Cambridge University Press, 1996.

[42] T. L. Griffiths, C. Kemp, and J. B. Tenenbaum, *Bayesian models of cognition*. Cambridge University Press, 2008.

[43] K. P. Körding and D. M. Wolpert, "Bayesian integration in sensorimotor learning," *Nature*, vol. 427, no. 6971, pp. 244–247, 2004.

[44] D. C. Van Essen *et al.*, "The WU-Minn human connectome project: an overview," *NeuroImage*, vol. 80, pp. 62–79, 2013.

[45] O. Y. Chén, "Book review of 'Big data in omics and imaging: Integrated analysis and causal inference'," *Journal of the American Statistical Association*, vol. 115, no. 529, pp. 487–488, 2020.

[46] ——, "The roles of statistics in human neuroscience," *Brain Sciences*, vol. 9, no. 8, p. 194, 2019.

[47] O. Y. Chén, R. G. Saraiva, H. Phan, J. Di, G. Nagels, T. Schwantje, H. Cao, J. Gou, J. M. Reinen, B. Xiong *et al.*, "The roles and challenges of the *P*-value," *arXiv*, p. 2002.07270, 2020.

[48] J. Fan and Y. Liao, "Endogeneity in high dimensions," *Annals of Statistics*, vol. 42, no. 3, p. 872, 2014.

[49] P. J. Huber, "Projection pursuit," *The Annals of Statistics*, vol. 13, no. 2, pp. 435–475, 1985.

[50] M. MacKay, H. B. Yang, S. M. Dursun, and G. B. Baker, "The gut-brain axis and the microbiome in anxiety disorders, post-traumatic stress disorder and obsessive-compulsive disorder," *Current Neuropharmacology*, vol. 21, pp. 1–18, 2023.

[51] N. Kumar, N. K. Sahoo, S. Mehan *et al.*, "The importance of gut-brain axis and use of probiotics as a treatment strategy for multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 71, p. 104547, 2023.

[52] L. Dohnalová, P. Lundgren, J. R. Carty, N. Goldstein, S. L. Wenski, P. Nanudorn, S. Thiengmag, K.-P. Huang, L. Litichevskiy, H. C. Descamps *et al.*, "A microbiome-dependent gut–brain pathway regulates motivation for exercise," *Nature*, vol. 612, pp. 739–747, 2022.