

Of networked lesions and lesioned networks

This scientific commentary refers to ‘Post-stroke outcomes predicted from multivariate lesion-behaviour and lesion network mapping’ by Bowren et al. (<https://doi.org/10.1093/brain/awac010>).

The excitement that greets reports of ‘convergent evidence’, perhaps greater in neuroscience than elsewhere, betrays their disturbing rarity. Each investigational modality throws the brain into a relief peculiarly its own, the areas of unqualified agreement so limited it is sometimes hard not to wonder if we are indeed looking at the same thing. The contrasting perspectives of correlative functional mapping in health, and disruptive deficit mapping in disease, present a striking example. The former, dominated by topological inference, paints an Elysian field of finely wrought, harmonious, replicable order; the latter, dominated by behavioural prediction, all too commonly offers crude, disjointed, shifting sketches of infernal chaos. Is either right, or none? Or is there a vantage point from which the two views become one? This is the implicit question Bowren¹ and his colleagues address in this issue of *Brain* in their elegant study of long-term stroke outcome prediction based on multivariate models that combine correlative data from normative populations—structural and functional connectivity maps—with disruptive data from individual patients—deficits and lesion masks.

Examining three large cohorts of structurally imaged patients with stroke, evaluated across an array of motor and cognitive outcomes over a 12-month period, the authors investigate the comparative fidelity of incrementally complicated predictive models incorporating acute behavioural data, multivariate lesion-deficit maps, and disruption to normative critical structural and functional connections. Quantified on the third, held out, cohort, they reassuringly find that lesion information adds long term predictive value to acute behaviour. Lesion-weighted compressed representations of critical white matter or resting state functional connections are shown to be superior to bare lesion-deficit maps on some, if not all outcomes, exhibiting domain-specific differences in the most informative connective modality. In short, contextualizing lesions by effects on connectivity inferred from normal anatomy and function is shown in some circumstances to enhance their predictive power.

Critical connectivity is here established by seeding normative data from the centroids of multivariate lesion-deficit maps that should already model spatially distributed dependencies. One can think of this as a deterministic projection of lesion features into a higher dimensional space where the outcome discriminative boundaries are thereby helpfully sharpened, a kind of ‘anatomical support vector machine’. That multivariate predictive models need it, even with substantial data, shows how hard the problem is, and how careful the model formulation must be.

We should distinguish the factors that unavoidably corrupt any model of the brain from those open to experimental remedy.

A deficit will always be under- or mis-specified behaviourally and physiologically: the former because human abilities are definitionally irreducible to scores, the latter because no instrument could conceivably capture every relevant neural process, disrupted or healthy. Equally, as far as a given brain instantiates a unique solution, it will be closed to understanding derived from observing the solutions of others. And as anyone who has trained an artificial deep neural network will know, the patterns of neural activation versus ablation sensitivity may vary radically in their coherence, not because the net has excess capacity but because its operation is constitutionally distributed, inhabiting a parameter space whose high dimensionality may take it far beyond the threshold of intelligibility. The real brain need be no different.

Our task is not to dwell on the insoluble problems but to stratify the soluble in order of their impact on model fidelity. Paramount here is proper recognition of the domain to which brain modelling belongs: like it or not, this is the study of complex systems. As proven at the dawn of cybernetics, an effective controller of a system can be no simpler than the system itself.² If by understanding is meant knowledge capable of guiding interventions—the principal concern of the translational neuroscientist—we must try to absorb complexity rather than reject it out of hand in the way simple models of the brain more or less openly do. This implies insisting on data scales as close to the ceiling of feasibility as we can get, evaluating mathematical models of the highest tractable expressivity, and deploying computation of commensurate power. It implies prioritizing the derivation of compact, yet rich, descriptions of the brain through representation learning, assigning primary importance to generalisable, individual-level predictive fidelity, and submitting to model comparison as the highest arbiter of credibility: policies still rare in a field dominated by criterial *P*-values drawn from simple inferential models whose statistical assumptions the data clearly violate.

Such recognition further compels us to relax implausible prior beliefs about the homogeneity of the brain across the population. Attention to the heterogeneous structure of lesions has already revealed its distorting effects not only on lesion-deficit maps,³ but also on the historical taxonomy of brain function itself.⁴ It is time we attended to the heterogeneity of the neural substrate too: the presence of systematic variability in its anatomical form. It is a foundational assumption, not only of mass-univariate inference but also patch-wise multivariate approaches applied at the group level, that anatomically identical regions support identical functions across the population. Where either the neural substrate or its anatomical landmarks vary in a manner image registration cannot correct, such models are bound to lose sensitivity in proportion to the departure from homogeneity. Note the substrate may still be structured, but now in terms of systematic subpopulation-specific

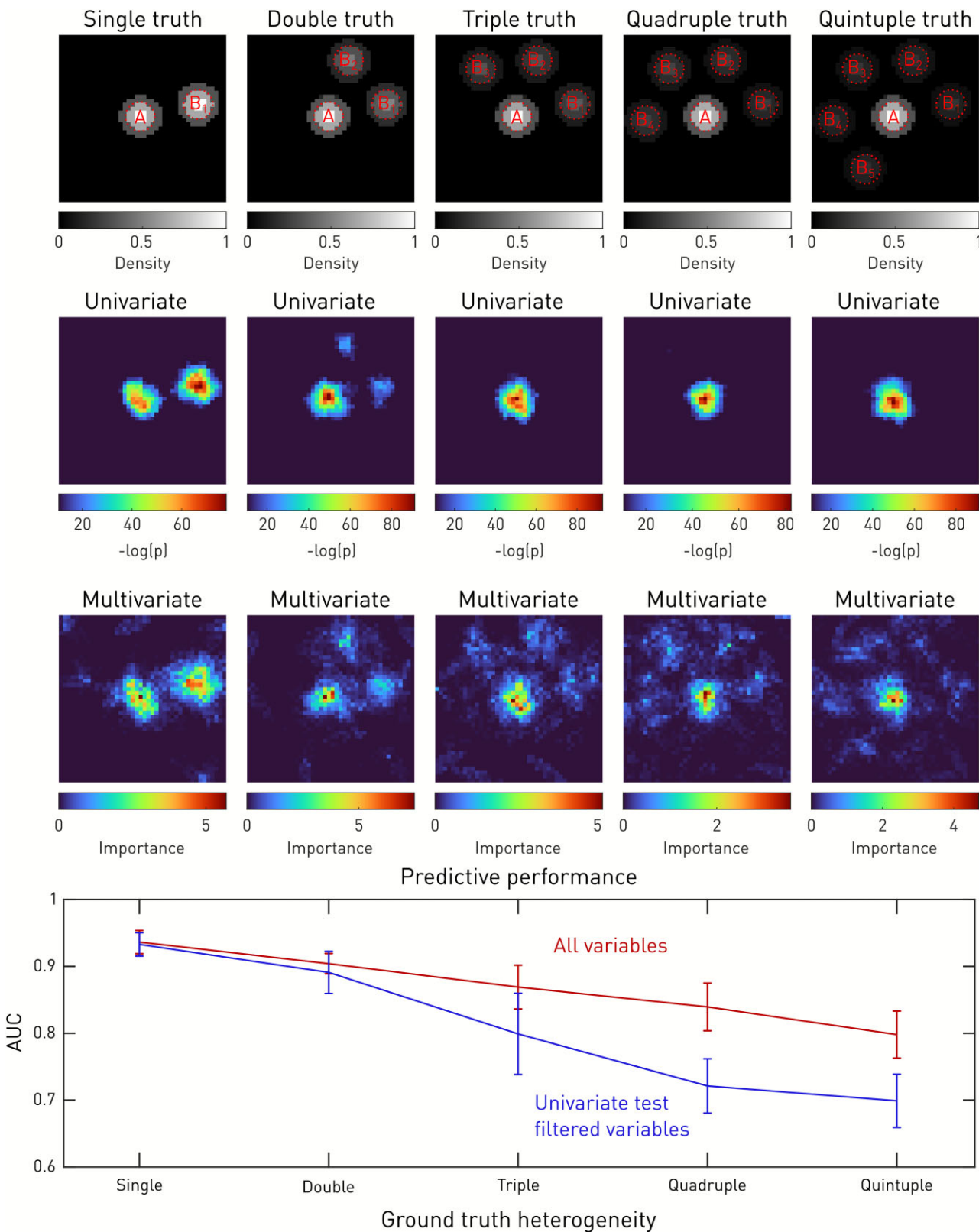


Figure 1 Demonstration with a planar (42x42 variable) synthetic model of the relation between lesion-deficit spatial inference and outcome prediction in the setting of heterogeneous neural substrates. A pair of areas, A and B, modelled as thresholded Gaussian blobs of randomly chosen size (first row), vary incrementally in the heterogeneity of their spatial distribution, from a fixed pattern (single pair), to five distinct patterns (quintuple pairs), across separate simulations. Region A is locationally fixed, but B has up to five possible locations B_1 to B_5 corresponding to increasing degrees of heterogeneity. A deficit generating process based on 10% overlap of a lesion with any part of A or B is used to create a synthetic ground truth from 1000 single, uniformly distributed synthetic lesions in the form of binary ellipses with a major to minor diameter ratio ranging between 0.2 and 0.3,

(Continued)

interactions between regions with variable localization. Such structure will be wholly opaque to simple methods, yet—given enough examples—potentially accessible to more flexible model architectures.

To illustrate the point, consider a hypothetical two-dimensional brain composed of 42×42 elements, where the critical substrate of a given function is distributed across two different regions: one, A, variable in size but fixed in location and another, B, variable in size and heterogeneous in location to an extent we manipulate across five hypothetical ground truth scenarios (Fig. 1, first row). If we attempt to map the substrate with a set of 1000 uniformly distributed ellipsoid ‘lesions’ of randomly varying orientation, assuming a symptom generating process that yields a deficit with 99% probability whenever a lesion overlaps with at least 10% of any part of the substrate, we can easily recover the location-invariant region A, whether we use mass-univariate or flexible multivariate inference (rows 2 and 3). Region B, by contrast, rapidly disappears from the picture with increasing heterogeneity, though less abruptly from multivariate feature importance maps. Crucially, multivariate predictive performance on the outcome, evaluated out of sample, remains comparatively high, and is higher for predictive models that use all available variables compared with those confined to the subset univariate testing identifies as significant (row 4).

In short, population heterogeneity in the neural substrate may be accessible to the right kind of model, but not in a way that can be represented by a simple anatomical map. This has two far-reaching implications. First, insisting on the use of predictive models explicable in terms of linearly composited voxel weights both imposes an unnecessarily low ceiling on maximum achievable fidelity and introduces potential bias against anatomically variable substrates. Seeking to avoid the misdemeanour of model opacity here risks committing the felony of poor and inequitable predictive performance. Second, heterogeneous substrates should be conceived as families of regional interactions more consistent in their topological structure than in the anatomical localization of their constituents: a form arguably best modelled as a hierarchically organized graph. Note the challenge is not modelling the

interactions between loci common across the population—a task on which graphs are already widely deployed—but characterizing subpopulation-specific structure primarily conveyed in network topologies. We need to become accustomed to looking less at technicolour pictures of the brain than at complex network communities describing multiple interacting regional effects. Our maps will be less recognizably of the brain, but all the truer for it.

Robert Gray and  Parashkev Nachev

UCL Queen Square Institute of Neurology, UCL, Queen Square, London, UK

Correspondence to: Parashkev Nachev

E-mail: p.nachev@ucl.ac.uk

Funding

The authors are funded by Wellcome (213038) and the NIHR UCLH Biomedical Research Centre.

Competing interests

The authors report no competing interests.

References

1. Bowren M, Bruss J, Manzel K, et al. Post-stroke outcomes predicted from multivariate lesion-behaviour and lesion network mapping. *Brain*. 2022;145:1338–1353.
2. Ashby WR. Requisite variety and its implications for the control of complex systems. In: *Facets of systems science*. Springer; 1991:405–417.
3. Mah YH, Husain M, Rees G, Nachev P. Human brain lesion-deficit inference remapped. *Brain*. 2014;137(Pt 9):2522–2531.
4. de Schotten MT, Foulon C, Nachev P. Brain disconnections link structural connectivity with function and behaviour. *Nat Commun*. 2020;11(1):5094.

Figure 1 Continued

yielding a deficit label vector. Fisher’s exact test was then applied in a pixel-wise manner to generate a field of asymptotic *P*-values, shown as the negative log in the second row, thresholded at a Bonferroni-corrected alpha of 0.05. The third row shows the feature importance from a high-dimensional multivariate model based on a sampling/boosting tree-based ensemble algorithm applied to the same data. At the bottom is a plot of out-of-sample predictive performance, indexed by the area under the receiver operating curve (AUC), with error bars ± 1 standard deviation from 10 independent runs, from the foregoing models (red) and another set of identically configured models limited to pixels identified as significant in the corresponding univariate model (blue). Note the differences in the ability of univariate versus multivariate models to retrieve the underlying substrate as heterogeneity increases, and the comparative predictive power of models with inputs constrained or not constrained to linearly associated features.