The sloppy relationship between neural circuit structure and function

Citation for published version:
Hennig, MH 2022, 'The sloppy relationship between neural circuit structure and function', *The Journal of Physiology*. https://doi.org/10.1113/JP282757

Digital Object Identifier (DOI):
10.1113/JP282757

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
The Journal of Physiology

TOPICAL REVIEW

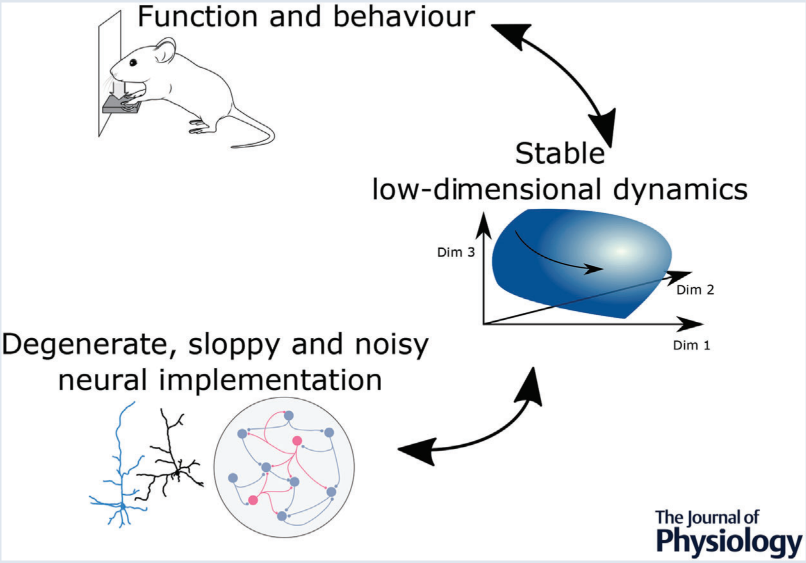
The sloppy relationship between neural circuit structure and function

Matthias H. Hennig

Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, Scotland

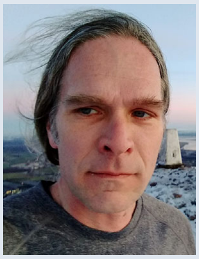
Handling Editors: Katalin Toth & Michael Okun

The peer review history is available in the Supporting information section of this article (https://doi.org/10.1113/JP282757#support-information-section).



Abstract Investigating and describing the relationships between the structure of a circuit and its function has a long tradition in neuroscience. Since neural circuits acquire their structure through sophisticated developmental programmes, and memories and experiences are maintained through synaptic modification, it is to be expected that structure is closely linked to function. Recent findings challenge this hypothesis from three different angles: function does not strongly constrain circuit parameters, many parameters in neural circuits are irrelevant and contribute little to function, and circuit parameters are unstable and subject to constant random drift. At the same time, however, recent work also showed that dynamics in neural circuit activity

Matthias Hennig is a Reader in Computational Neuroscience at the School of Informatics, University of Edinburgh. He studied Physics at Bochum University and then obtained a PhD in Computational Neuroscience at the University of Stirling in 2005. His research is focused on development, stability and processing in neuronal circuits using computational modelling. To enable links between theory and experiments, his group also develops methods and open-source tools for analysis and interpretation of large scale neural recordings, and develops and advocates open science approaches.



DOI: 10.1113/JP282757

that is related to function are robust over time and across individuals. Here this apparent contradiction is addressed by considering the properties of neural manifolds that restrict circuit activity to functionally relevant subspaces, and it will be suggested that degenerate, anisotropic and unstable parameter spaces are closely related to the structure and implementation of functionally relevant neural manifolds.

(Received 14 April 2022; accepted after revision 20 July 2022; first published online 25 July 2022)
Corresponding author M. H. Hennig: Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, Scotland. Email: m.hennig@ed.ac.uk

Abstract figure legend What are the relationships between noisy and highly variable microscopic neural circuit variables on the one hand and the generation of behaviour on the other? Here it is proposed that an intermediate level of description exists where this relationship can be understood in terms of low-dimensional dynamics. Recordings of neural activity during unconstrained behaviour and the development of new machine learning methods will help to uncover these links.

Introduction

The seminal work by Hubel and Wiesel published in this journal (Hubel & Wiesel, 1962) proposed a hierarchical circuit model for orientation selectivity in the visual cortex. In this model, the selectivity of simple and complex cells follows from convergent connectivity. While it was only a hypothesised circuit at this point, it gave rise to the idea that the function of neural circuits could be understood in terms of their connectivity. Indeed central predictions of the Hubel and Wiesel model were subsequently confirmed in combined physiological and tracing studies (Gilbert, 1983; Gilbert & Wiesel, 1983), which also revealed a highly structured organisation of neurons and their connectivity in the different cortical layers. This, in turn, led to the hypothesis that cortical circuits are composed of small, functional microcircuits that are flexibly combined to give rise to a variety of functions (Douglas & Martin, 2004; Nelson, 2002). In analogy with electronics circuits, research in the second half of 20th century was influenced by the view that understanding brain function required a precise wiring diagram of the neurons, and that knowing this wiring diagram would directly explain function – one would just have to turn the circuit on.

These influential findings still resonate today, and contemporary research in systems neuroscience is often explicitly or implicitly based on the premise that there is a firm and interpretable relationship between the function of a neural circuit and its synaptic connectivity (Yuste, 2015). While the complexity and diversity of neural excitability and synaptic function that complicates direct structure–function inference are of course generally appreciated (Morgan & Lichtman, 2013), this thinking motivated detailed high-throughput connectomics studies that now provide valuable data sets to complement physiological and theoretical studies (Bae et al., 2021; Cook et al., 2019; Milyaev et al., 2011).

Yet recent findings show that structure–function relationship might not be as tight as once thought: the parameter landscapes of circuits can be highly degenerate, many parameters appear only weakly constrained by function, and circuit activity is unstable and slowly drifts over time. This includes not only synaptic connections, but also the full ion channel complement of each neuron, which determines its excitability and, of course, the mechanism for action potential generation celebrated in this issue (Hodgkin & Huxley, 1952). In parallel, as recording technologies rapidly advance, there has been a shift of focus from single neurons to the activity of large populations, which reveals complex distributed dynamics with unclear relationships between function and circuit structure and neural physiology. This review explores these seemingly disparate findings and offers a re-interpretation in the context of recent results studying neural population dynamics.

Characterising circuit function

To understand circuit function, we record and analyse neural activity and attempt to quantify how it relates to external variables such as stimuli or behaviour. In some cases, carefully designed experiments reveal a clear-cut and fully interpretable relationship between function and structure. One such example is the Reichardt detector circuit, an elementary motion detector in the fly retina where the activity of two photoreceptor inputs are filtered differently so that the order of their activation affects their combined response (Reichardt, 1987). Knowing that the primary visual cortex receives structured thalamic inputs enabled Hubel and Wiesel to describe its function as an early image-feature detector and to propose an underlying circuit.

To better understand the insights gained from such models, it is useful to consider David Marr's hierarchy of three levels of analysis for information processing systems. At the top of this hierarchy is the computational level which describes the function (or objective, for instance to generate an appropriate behaviour) of the system. Next, the algorithmic level captures the mathematical approach used to realise this function. At the bottom is the implementational or mechanistic level, the actual realisation of this algorithm with neurons. In the cases above, the mechanistic level is well understood and enabled descriptions at the algorithmic and computational levels. However, Hubel and Wiesel's model fails to predict the responses of simple cells to arbitrary stimuli, so it is likely incomplete and captures only one of many aspects of the operation of the primary visual cortex (David & Gallant, 2005). In fact, we now realise that the primary visual cortex is not only modulated by visual stimuli, but also by behaviour (Flossmann & Rochefort, 2021; Niell & Stryker, 2010), so our understanding at the computational level still seems incomplete.

These simple examples were chosen to make the broader point: reasoning about circuit function can be complex, and the question of whether a circuit performs a specific function may be ill posed without taking endogenous and exogenous states into account (Bassett & Gazzaniga, 2011). While we can attribute function to a circuit by examining which stimulus-related, cognitive or behavioural variables can be decoded from its activity, this usually cannot sufficiently constrain the algorithmic level of explanation. Equally, explanations at the computational level are difficult to infer from the algorithmic level if its explanation is incomplete. This may seem an obvious limitation of laboratory studies where the effects of stimuli or behaviour have to be analysed under constrained and controlled conditions. Advances in recording technologies and simultaneous behaviour monitoring may provide data sets that allow constructing models that generalise better across different contexts (Urai et al., 2022). So far, however, it is critical to remember that our definitions of circuit function may still be rather tentative and limited.

In the following therefore, we will use two practical yet somewhat flexible definitions of circuit function. One is to simply ask which external (and possibly also endogenous) variables modulate the recorded activity, for instance using a decoder. In this case the algorithmic or computational implications may be unclear, but their importance is implied by association. The second is to simply characterise the activity repertoire of the circuit, which includes the different observed firing patterns and their temporal order, without paying attention to other variables (such as the inputs a circuit receives), again assuming these statistics are functionally relevant through association. Using these working definitions of function, we will next discuss three recent findings in the context of circuit structure–function relationship.

Circuit parameter degeneracy

A first important result on the structure–function relationship in neural circuits is that the same behaviour can arise from very different parameter combinations (Fig. 1*A*). This means that neural circuits show highly degenerate parameter spaces. These parameters include synaptic and cell-intrinsic conductance which combine non-linearly to produce the circuit activity. Computational models have been an important tool to study such complex systems for a long time, and anyone who has worked with models will likely have encountered a situation where the available experimental data leave a model hopelessly under-constrained. This problem was systematically addressed by Golowasch et al. (2002) who compared the behaviour of single neurons to a collection of randomly generated models. This revealed that the same activity could be replicated with a whole family of models: often a particular physiological phenotype could be maintained by compensating a change in one conductance by a corresponding change of another. Importantly, the non-linear behaviour of many ion channels leads to a complex parameter landscape where averages are uninformative: a circuit derived from averaging parameters from many experiments usually differs in behaviour from that of the individual specimen (Golowasch et al., 2002).

A series of elegant experimental and theoretical studies from Eve Marder's group has shown that neural systems indeed exploit this flexibility. Studying a circuit in stomatogastric ganglia in crustaceans where function is clearly defined by the rhythmic pattern it generates, they reported a high variability of conductances, synaptic and intrinsic, between animals whose pyloric circuits produced similar oscillations (Marder et al., 2015). These results suggest that neural circuits are constrained to produce a desired output or function, while the implementation of this function is flexible. This likely provides considerable flexibility and robustness as it enables the acquisition and maintenance of function through different developmental trajectories, environments and life-long homeostasis (O'Leary et al., 2014).

Sloppy parameter spaces

A second important phenomenon is that for a given circuit configuration a large number of directions in parameter space have little or no influence on circuit function, which is constrained by only a few very relevant directions (Fig. 1*B*). Importantly, these *directions* in parameter space are usually oriented along combinations of parameters and not the bare parameter axes, similar to how degeneracy arises from compensatory changes. The term 'sloppiness' refers to systems with such anisotropic sensitivity profiles where only a very small number of parameter space directions determine their behaviour. To show this, it is again necessary to first define the function of a circuit. This may be firing patterns such as the oscillations in the pyloric circuit, or also the precision of a read-out of encoded quantities such as a stimulus. To evaluate parameter sensitivity, we start with a set of experimentally determined parameters, and then systematically change each parameter slightly and record the change in output or function (in the limit of small perturbations). These results are then summarised in a Hessian matrix that quantifies the curvature of this error in all possible parameter directions (Transtrum et al., 2015). Analysing this matrix can reveal which circuit parameters are important and strongly constrained, and which ones can be changed without affecting function. Computational models are essential to perform this analysis as it is experimentally infeasible to systematically change many neuron and circuit parameters. Computing the Hessian matrix is typically hard, but for certain statistical model classes of functional circuit connectivity (for instance exponential family models), it can be obtained in closed form.

This method was applied to activity recorded from cultured networks with high density microelectrode arrays, and the importance of the functional connectivity parameters for maintaining the correlation structure of the recorded activity was evaluated (Panas et al., 2015). This showed not only that parameter changes can be compensated by changes in other parameters, but that the vast majority of parameters were irrelevant while only a small number of couplings determined the circuit activity. In other words, neural circuits have a hierarchy of directions in parameter space that are increasingly less relevant, and thus can change without consequence for a desired behaviour. This is a common property of biological models with large numbers of parameters (Gutenkunst et al., 2007). Interestingly, sensory stimulation alters activity in cortical networks along sloppy directions in functional connectivity space, while switching between synchronised and desynchronised cortical states occurs along the sensitive directions (Ponce-Alvarez et al., 2020). This suggests that the development of cortical networks is constrained such that modulatory processes can act to change the network state effectively.

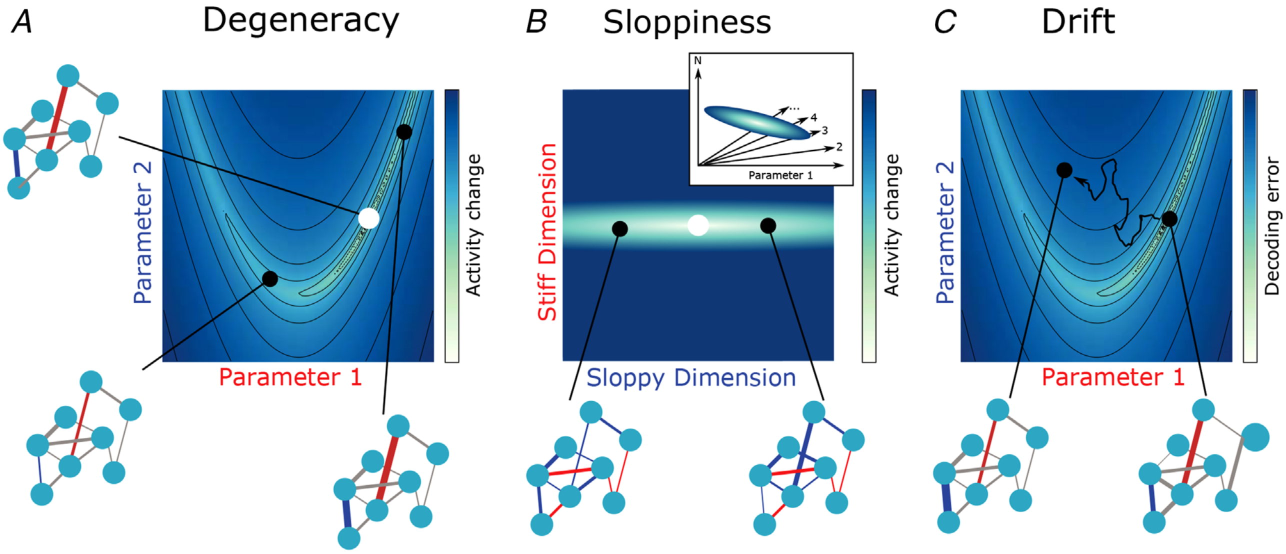


Figure 1. Three properties of neural circuits that complicate relating structure to function
A, in circuits with parameter degeneracy, the same activity is obtained for different parameter combinations. Activity is compared to a reference point (white circle) with an error measure to quantify the degree of deviation (activity change) from the reference circuit. In some cases, the change of one parameter can be compensated by changing a second, which leads to the bright area with small error in the figure. These regions of low error are often curved, and thus effective compensation for parameter changes is non-linear. *B*, sloppiness refers to anisotropic parameter spaces with few important and many irrelevant directions. After selecting a reference circuit (white circle), the curvature of the error measure is examined for changes along all parameters (see inset). In sloppy models, the contours of same sensitivity (same colour) are arranged on highly eccentric ellipses where the long axis points towards the least relevant parameter combination (see inset). In the main plot, these two axes are shown so that a stiff, sensitive dimension is vertically aligned and an insensitive, sloppy dimension horizontally. Note that these dimensions now denote parameter combinations (see network cartoons), not single parameters as in *A*. *C*, representational drift is the change of a circuit due to random fluctuations. Often this is based on decoding of an external variable such as sensory stimuli. The random movement in parameter space typically causes an increase in decoding error as non-optimal parameter regimes are reached.

Representational drift

Third, neural circuits are unstable and subject to continuous slow modifications that, over days and weeks, can degrade stored information and function (Fig. 1*C*). Early studies investigating the decoding of movement from the motor cortex reported some variability in the activity of neurons over time, yet surprisingly little concomitant degradation in decoding performance (Carmena et al., 2005; Chestek et al., 2007). In these electrophysiological experiments the lack of stability of the probe can be a contributing factor, but more recent experiments using stable long-term imaging have shown that neural representations indeed change slowly over time (Driscoll et al., 2017; Huber et al., 2012; Ziv et al., 2013). This phenomenon, termed 'representational drift', appears to be ubiquitous in the cortex (Deitch et al., 2021; Marks & Goard, 2021; Rubin et al., 2015; Schoonover et al., 2021). It is important to note, however, that the functional implications of representational drift are still unclear as its effect on behaviour was not assessed in most of these experiments. Driscoll et al. (2017) recorded neurons in superficial layers of the posterior parietal cortex (PPC) in mice trained in a T-maze task and reported stable behaviour performance in the presence of substantial drift. While PPC is required in this task and the results are suggestive as the neurons respond in a highly task-related manner, it is still difficult to know whether the recorded neurons causally contribute to task performance. This is pertinent as some studies also report stable neural representations in motor areas of song birds (Katlowitz et al., 2018) and place cells in rats trained in navigation tasks (Wirtshafter & Disterhoft, 2022).

The precise origin of this ongoing circuit re-modelling is debated, but it likely includes stochastic processes due to continuous synapse proteome turnover (Maletic-Savatic et al., 1999; Minerbi et al., 2009; Okabe et al., 1999; Raman & O'Leary, 2021) that leads to changes in synaptic strength and causes synapse turnover (Mongillo et al., 2017), and ongoing experience-dependent plasticity (Holtmaat et al., 2005; Trachtenberg et al., 2002). Equally, neural excitability is subject to change and thus likely also contributes to the observed variability (Daoudal & Debanne, 2003; Desai et al., 1999) although little is known about the contribution of these mechanisms to representational drift. Recent theoretical work has shown that such volatility may not be functionally disruptive as simple and plausible plasticity rules can effectively compensate for slow changes (Raman & O'Leary, 2021; Rule & O'Leary, 2022; Rule, Loback et al., 2020). Instead, it has been suggested that representational drift may be beneficial for more effective and flexible learning (Rule et al., 2019).

The structure of neural manifolds

The three phenomena discussed above suggest it is challenging to establish a tight and generalisable relationship between circuit function and structure. However, population activity is highly structured and often confined to low-dimensional manifolds in the high-dimensional ambient space spanned by the activity of each neuron (Cunningham & Byron, 2014). A neural manifold is a locally connected geometrical subspace that resides in a high-dimensional space and can be viewed as the collection of the accessible repertoire of activity patterns of the circuit. Activity moves along this manifold as neurons change their activity. A simple example is a ring attractor circuit, used to model orientation selective neurons and head direction cells. Here the combination of short range excitation and long-range inhibition constrains the activity to be on ring, and thus on a one-dimensional manifold (Amari, 1977; Ben-Yishai et al., 1995; Skaggs et al., 1994). A more complex manifold was recently characterised in grid cell populations of the medial entorhinal cortex (Gardner et al., 2022), which has a doughnut-shaped geometry as predicted by theoretical work (Chaudhuri et al., 2019).

To uncover low-dimensional manifolds from population activity, latent variable models can be employed to summarise the joint activity in a small number of dimensions (Hurwitz, Kudryashova et al., 2021). Early applications of such methods include the discovery of low-dimensional population dynamics in the olfactory and motor system (Churchland et al., 2012; Mazor & Laurent, 2005), which sparked the development of various machine learning approaches for the analysis of population activity (reviewed by Hurwitz, Kudryashova et al., 2021).

An important insight from this work is that circuit activity not only evolves through the inputs to the circuit, but depends strongly on its recurrent connections and on recurrent interactions between brain areas. For instance, movement-related activity in the motor cortex is well-described through dynamics that require only an input prior to movement onset. Once set in motion, activity subsequently evolves autonomously through the recurrent connections without need for further inputs (Hurwitz, Srivastava et al., 2021; Sussillo et al., 2016). In this context, different movements are represented by different trajectories, and together they form a complex manifold to enable flexible control of movements. As a result, the recurrent connectivity not only contributes to but may in fact dominate circuit activity, and the role of inputs is to control or to modulate the movement on this manifold.

In this scenario the role of circuit connectivity is, in concert with neuronal physiology, to sculpt out relevant manifolds on which circuit dynamics evolve. This link is, however, tenuous. There is usually no unique mapping between dynamics and connectivity since the same low-dimensional activity can be created by many possible, disparate networks: circuit degeneracy also applies to manifolds. Hence it is important to ask whether there are constraining factors that shape neural manifolds, or if the brain indeed starts out as a *tabula rasa* where manifolds are shaped by developmental history and experience as appears to be the case in the pyloric rhythm circuit.

Simultaneous assessment of the selectivity of neurons in the superficial layers of the visual cortex and their synaptic connections has shown that neurons with similar tuning are more likely recurrently connected by strong synapses (Cossell et al., 2015). This indicates the presence of strong constraints during circuit formation with clear functional implications, so the structure of the networks that emerge during circuit development is constrained to a subset of the possible functional networks. Various mechanisms may be responsible, including biochemical guidance molecules, anatomical constraints and variations in plasticity rules. At the same time, the majority of recurrent synapses in V1 are weak and show no clear preferences in their connectivity, so appear only weakly constrained.

A hint as to how neural manifolds are built comes from the analysis of the sloppy and relevant or ‘stiff’ directions in the circuit parameter space. Panas et al. (2015) reported that especially neurons with high firing rates define the stiff directions in functional connectivity models. This was confirmed in a study of simulated circuits with realistic connectivity parameters, where random permutations of synapses has little effect on the population activity for excitatory synapses, but a strong effect for inhibitory synapses (Mongillo et al., 2018). A recent study extends these insights by studying network models of working memory trained to retain information (Kim & Sejnowski, 2021). This work not only confirms the specific importance of inhibitory neurons, but also identifies a specific circuit motif of mutual inhibition between different inhibitory populations in these otherwise highly variable networks. To provide a further example, specific inhibitory microcircuits have been proposed as a critical element for the temporal integration of feed-forward and top-down inputs (Wilmes & Clopath, 2019).

Together these results suggest that potentially tight relationships between circuit structure and function are to be found in inhibitory circuits (Herstel & Wierenga, 2021) that form a ‘backbone’ around which network activity evolves flexibly (Buzsáki & Mizuseki, 2014). Why then do neural circuits have large unconstrained regions in their parameter spaces, in particular in the excitatory populations? One possible explanation relates to the capacity of the circuit. Theoretical work has shown that networks trained to discover and learn the structure of high-dimensional inputs such as sensory information enter sloppy (or critical) regimes when their capacity is sufficient to optimally encode and compress (Cubero et al., 2019; Marsili & Roudi, 2022; Rule, Sorbaro et al., 2020). This predicts that the excitatory connections in a circuit, taken together, hold information, but the weight of each single one is significantly less important than an inhibitory weight. Could such sloppy connections simply be removed? Indeed, a theoretical study suggests that neurons can locally compute the importance of a synapse, which allows pruning of connections and neurons with minimal impact on circuit function (Scholl et al., 2021). This is a potential model for developmental pruning during which significant numbers of synapses and neurons are removed. Yet pruning always impacts the capacity of a network, and so involves a trade-off between cost and function.

Finally, representational drift should also cause instabilities in neural manifolds. Surprisingly this seems not the case (Fig. 2). In a recent study that analysed population activity from the primate motor cortex recorded over many months, movement-related latent neural dynamics were found to be stable for as long as 2 years (Gallego et al., 2020). With non-linear decoders and a model that implements dynamics on a stable manifold, it is possible to correctly predict behaviour after a simple alignment procedure (Dabagia et al., 2020; Farshchian et al., 2018; Wen et al., 2021), or even without alignment (Jude et al., 2022). Moreover, neural manifolds underlying odour encoding can be aligned across animals using a surprisingly simple linear approach, indicating that manifold development and structure are strongly constrained (Herrero-Vidal et al., 2021). Stability over time has also been documented in neural populations storing associative memories. Neurons co-active during conditioning form a memory engram and are not only reactivated by the conditioned stimulus (e.g. a tone), but their optogenetic activation also elicits a conditioned response (e.g. freezing in anticipation of a foot shock). Critically, this reactivation has been reported for up to 2 weeks after training, a duration in which considerable representational drift would be expected (Josselyn & Tonegawa, 2020).

There are several, not mutually exclusive, mechanisms that could underlie the robustness of neural manifolds and of specific functional connectivity more generally. A first is that neural circuits can compensate for drift by continuously updating weights. Models demonstrate that simple Hebbian and homeostatic mechanisms can effectively adapt read-out weights (Rule & O’Leary, 2022) or the weights of the entire system (Kossio et al., 2021) such that stable function is preserved. This offers considerable flexibility to circuits to continuously re-organise and adapt, but seems at odds with stable manifolds. Alternatively, plasticity can actively compensate for drifts by selectively maintaining the relevant connections. For instance, the reactivation of sequences on the manifold through spontaneous activation can strengthen relevant connections (Kossio et al., 2021), and a recent modelling study demonstrated that such replay can indeed protect associative memories against drift (Fauth & van Rossum, 2019). Theoretical analyses show that generally compensatory plasticity is effective as long as its magnitude matches that of the drift, and so it can be effective even when there is only occasional re-activation (Raman & O'Leary, 2021). Finally, the analysis of the population activity in neural cultures and the visual cortex shows that drift is not uniform, but occurs predominantly along the sloppy parameter directions (Panas et al., 2015; Sweeney & Clopath, 2020). Since the sloppy part of the parameter space is typically large, many neurons will change their activity in this scenario and one may conclude the population drifts as a whole. However, at the same time the correlation structure of the activity is preserved such that the population read-out can remain stable as long as it is well aligned to the stable neural manifold. The latter two mechanisms predict that at least some aspects of the ensemble activity will remain stable during representational drift. Recent machine learning approaches to discover behaviourally relevant manifolds will make it possible to ask directly if this weak stability requirement at the circuit level is sufficient for stable behaviour performance (Hurwitz, Srivastava et al., 2021; Sani et al., 2020).

Conclusions and outlook

This review started with a suggestion that degeneracy, sloppiness and spontaneous drift challenge the notion of well-defined relationships between structure and function in neural circuits. However, analysing these ubiquitous circuit properties in the context of neural manifolds suggests a different explanation. Instead, these properties may reflect organising principles of high-dimensional, highly structured dynamical systems. Sloppiness is an important property as this indicates that simplification and explanation is achievable through a procedure known as coarse graining, which aims to compress a complex system into simpler ones by removing sloppy degrees of freedom and summarising stiff variables (Machta et al., 2013).

Coarse graining is precisely what latent variable models aim to achieve by describing a high dimensional system in terms of a small number of causal factors. Therefore, understanding sloppiness in neural circuits can help understanding their function. Why then is sloppiness so ubiquitous in biological systems, what is its function? An analysis of systems biology models suggests sloppy systems can achieve both high robustness and evolvability (Daniels et al., 2008), which are desired properties also in the brain. Sloppy parameter spaces can also be a signature of an optimal information encoder and other, optimally functioning, learning systems (Cubero et al., 2019; Marsili & Roudi, 2022; Rule, Sorbaro et al., 2020). However, these findings still have to be connected more precisely to neural circuit function, and in particular to circuit dynamics.

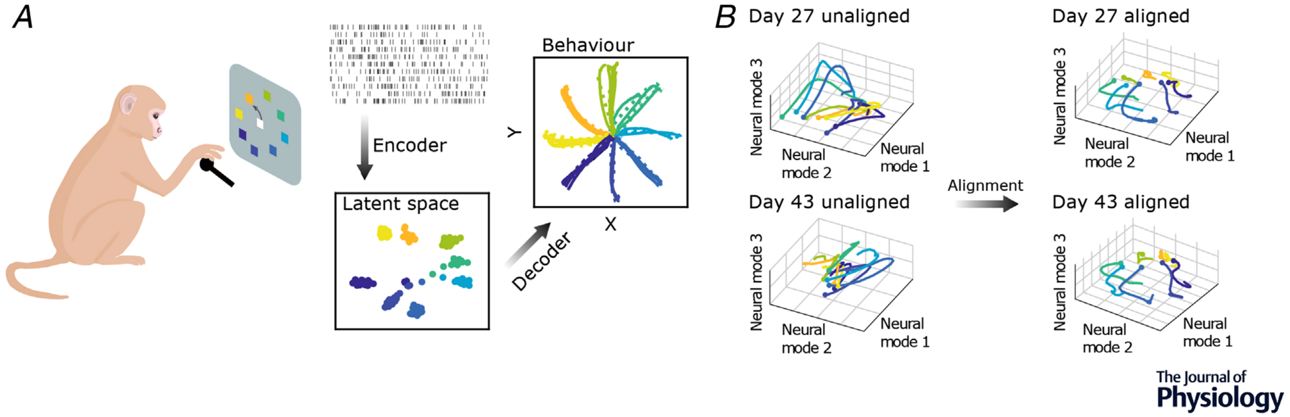


Figure 2. Latent variable models characterise neural manifolds and uncover long-term stability in latent neural dynamics
A, population activity recorded from the primate primary motor cortex during a centre-out reach task has a low-dimensional, behaviour-related latent embedding. The latent space shown here was extracted using TNDM, a non-linear state space model that extracts behaviourally relevant dynamics from neural activity (Hurwitz, Srivastava et al., 2021). The latent dynamics (not illustrated) allow precise behaviour decoding. *B*, despite the significant changes in recorded activity due to drift and variability in extracellular recordings, it is possible to align activity recorded many days apart from the same animal into a common, stable latent space. This shows that despite various sources of instability, the circuit components that sculpt out these latent dynamics are stable over long periods of time. The figure was modified from Gallego et al. (2020).

It will therefore be fruitful for future work to combine bottom-up and top-down approaches as it is plausible that both can converge at an intermediate level that perhaps corresponds to Marr's algorithmic level. Promising developments that will help achieve this include the rapid development of technologies to monitor large populations during unconstrained behaviour, and the development of advanced machine learning tools to analyse complex behaviour and neural recordings. Such experiments can potentially reveal important missing links between the levels of explanation: behaviour defines the computational level as it reflects the various objectives the brain solves while neural recordings simultaneously provide a window into the mechanistic implementation. Interpretable latent variable models that correctly predict behaviour from neural activity can then be analysed to reveal potential hypotheses at the algorithmic level, which in turn can yield testable predictions to reject or refine them, and to test their generality. To achieve this, work and collaboration is required in many domains: we need carefully crafted experimental designs, improved and scalable analysis and modelling methodology, and to pay close attention to reproducibility in the face of complexity and high variability. After all, attempting to bridge different levels of explanations has been fruitful already in the days of Hodgkin and Huxley and will likely continue to be so.

References

Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, **27**(2), 77–87.

Bassett, D. S., & Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends in Cognitive Sciences*, **15**(5), 200–209.

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences, USA*, **92**(9), 3844–3848.

Buzsáki, G., & Mizuseki, K. (2014). The log-dynamic brain: How skewed distributions affect network operations. *Nature Reviews. Neuroscience*, **15**(4), 264–278.

Carmena, J. M., Lebedev, M. A., Henriquez, C. S., & Nicolelis, M. A. (2005). Stable ensemble performance with single-neuron variability during reaching movements in primates. *Journal of Neuroscience*, **25**(46), 10712–10716.

Chaudhuri, R., Gerçek, B., Pandey, B., Peyrache, A., & Fiete, I. (2019). The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, **22**(9), 1512–1520.

Chestek, C. A., Batista, A. P., Santhanam, G., Byron, M. Y., Afshar, A., Cunningham, J. P., Gilja, V., Ryu, S. I., Churchland, M. M., & Shenoy, K. V. (2007). Single-neuron stability during repeated reaching in macaque premotor cortex. *Journal of Neuroscience*, **27**(40), 10742–10750.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, **487**(7405), 51–56.

Bae, J. A., Baptiste, M., Bodor, A. L., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., Celii, B., Cobos, E., Collman, F., da Costa, N. M., Dorkenwald, S., Elabbady, L., Fahey, P. G., Fliss, T., Froudarakis, E., Gager, J., Gamlin, C., …, Yu, S., MICrONS Consortium (2021). Functional connectomics spanning multiple areas of mouse visual cortex. *bioRxiv*. https://doi.org/10.1101/2021.07.28.454025. bioRxiv

Cook, S. J., Jarrell, T. A., Brittin, C. A., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., Nguyen, K. C., Tang, L. T.-H., Bayer, E. A., Duerr, J. S., Bülow, H. E., Hobert, O., Hall, D. H., & Emmons, S. W. (2019). Whole-animal connectomes of both caenorhabditis elegans sexes. *Nature*, **571**(7763), 63–71.

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, **518**(7539), 399–403.

Cubero, R. J., Jo, J., Marsili, M., Roudi, Y., & Song, J. (2019). Statistical criticality arises in most informative representations. *Journal of Statistical Mechanics: Theory and Experiment*, **2019**(6), 063402.

Cunningham, J. P., & Byron, M. Y. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, **17**(11), 1500–1509.

Dabagia, M., Kording, K. P., & Dyer, E. L. (2020). Comparing high-dimensional neural recordings by aligning their low-dimensional latent representations. *arXiv preprint arXiv:220508413*.

Daniels, B. C., Chen, Y. J., Sethna, J. P., Gutenkunst, R. N., & Myers, C. R. (2008). Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, **19**(4), 389–395.

Daoudal, G., & Debanne, D. (2003). Long-term plasticity of intrinsic excitability: Learning rules and mechanisms. *Learning & Memory*, **10**, 456–465.

David, S. V., & Gallant, J. L. (2005). Predicting neuronal responses during natural vision. *Network: Computation in Neural Systems*, **16**(2–3), 239–260.

Deitch, D., Rubin, A., & Ziv, Y. (2021). Representational drift in the mouse visual cortex. *Current Biology*, **31**(19), 4327–4339.e6.

Desai, N. S., Rutherford, L. C., & Turrigiano, G. G. (1999). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neuroscience*, **2**(6), 515–520.

Douglas, R. J., & Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, **27**(1), 419–451.

Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N., & Harvey, C. D. (2017). Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, **170**(5), 986–999.e16.

Farshchian, A., Gallego, J. A., Cohen, J. P., Bengio, Y., Miller, L. E., & Solla, S. A. (2018). Adversarial domain adaptation for stable brain-machine interfaces. *arXiv preprint arXiv:181000045*.

Fauth, M. J., & van Rossum, M. C. (2019). Self-organized reactivation maintains and reinforces memories despite synaptic turnover. *eLife*, **8**, e43717.

Flossmann, T., & Rochefort, N. L. (2021). Spatial navigation signals in rodent visual cortex. *Current Opinion in Neurobiology*, **67**, 163–173.

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., & Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, **23**(2), 260–270.

Gardner, R. J., Hermansen, E., Pachitariu, M., Burak, Y., Baas, N. A., Dunn, B. A., Moser, M.-B., & Moser, E. I. (2022). Toroidal topology of population activity in grid cells. *Nature*, **602**(7895), 123–128.

Gilbert, C. D. (1983). Microcircuitry of the visual cortex. *Annual Review of Neuroscience*, **6**(1), 217–247.

Gilbert, C. D., & Wiesel, T. N. (1983). Functional organization of the visual cortex. *Progress in Brain Research*, **58**, 209–218.

Golowasch, J., Goldman, M. S., Abbott, L., & Marder, E. (2002). Failure of averaging in the construction of a conductance-based neuron model. *Journal of Neurophysiology*, **87**(2), 1129–1131.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*, **3**(10), e189.

Herrero-Vidal, P., Rinberg, D., & Savin, C. (2021). Across-animal odor decoding by probabilistic manifold alignment. *Advances in Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/2021/file/aad64398a969ec3186800d412fa7ab31-Paper.pdf

Herstel, L. J., & Wierenga, C. J. (2021). Network control through coordinated inhibition. *Current Opinion in Neurobiology*, **67**, 34–41.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, **117**(4), 500–544.

Holtmaat, A. J., Trachtenberg, J. T., Wilbrecht, L., Shepherd, G. M., Zhang, X., Knott, G. W., & Svoboda, K. (2005). Transient and persistent dendritic spines in the neocortex in vivo. *Neuron*, **45**(2), 279–291.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**(1), 106–154.

Huber, D., Gutnisky, D. A., Peron, S., O'Connor, D. H., Wiegert, J. S., Tian, L., Oertner, T. G., Looger, L. L., & Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature*, **484**(7395), 473–478.

Hurwitz, C., Kudryashova, N., Onken, A., & Hennig, M. (2021). Building population models for large-scale neural recordings: Opportunities and pitfalls. *Current Opinion in Neurobiology*, **70**, 64–73.

Hurwitz, C., Srivastava, A., Xu, K., Jude, J., Perich, M., Miller, L., & Hennig, M. H. (2021). Targeted neural dynamical modeling. *Advances in Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/2021/file/f5cfbc876972bd0d031c8abc37344c28-Paper.pdf

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, **367**(6473), eaaw4325.

Jude, J., Perich, M. G., Miller, L. E., & Hennig, M. H. (2022). Robust alignment of cross-session recordings of neural population activity by behaviour via unsupervised domain adaptation. *Proceedings of the 39th International Conference on Machine Learning*, PMLR 162, 10462-10475.

Katlowitz, K. A., Picardo, M. A., & Long, M. A. (2018). Stable sequential activity underlying the maintenance of a precisely executed skilled behavior. *Neuron*, **98**(6), 1133–1140.e3.

Kim, R., & Sejnowski, T. J. (2021). Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nature Neuroscience*, **24**(1), 129–139.

Kossio, Y. F. K., Goedeke, S., Klos, C., & Memmesheimer, R.-M. (2021). Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences, USA*, **118**(46), e2023832118.

Machta, B. B., Chachra, R., Transtrum, M. K., & Sethna, J. P. (2013). Parameter space compression underlies emergent theories and predictive models. *Science*, **342**(6158), 604–607.

Maletic-Savatic, M., Malinow, R., & Svoboda, K. (1999). Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. *Science*, **283**(5409), 1923–1927.

Marder, E., Goeritz, M. L., & Otopalik, A. G. (2015). Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Current Opinion in Neurobiology*, **31**, 156–163.

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature Communications*, **12**, 5169.

Marsili, M., & Roudi, Y. (2022). Quantifying relevance in learning and inference. *arXiv preprint arXiv:220200339*.

Mazor, O., & Laurent, G. (2005). Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons. *Neuron*, **48**(4), 661–673.

Milyaev, N., Osumi-Sutherland, D., Reeve, S., Burton, N., Baldock, R. A., & Armstrong, J. D. (2011). The virtual fly brain browser and query interface. *Bioinformatics*, **28**(3), 411–415.

Minerbi, A., Kahana, R., Goldfeld, L., Kaufman, M., Marom, S., & Ziv, N. E. (2009). Long-term relationships between synaptic tenacity, synaptic remodeling, and network activity. *PLoS Biology*, **7**(6), e1000136.

Mongillo, G., Rumpel, S., & Loewenstein, Y. (2017). Intrinsic volatility of synaptic connections—A challenge to the synaptic trace theory of memory. *Current Opinion in Neurobiology*, **46**, 7–13.

Mongillo, G., Rumpel, S., & Loewenstein, Y. (2018). Inhibitory connectivity defines the realm of excitatory plasticity. *Nature Neuroscience*, **21**(10), 1463–1470.

Morgan, J. L., & Lichtman, J. W. (2013). Why not connectomics? *Nature Methods*, **10**(6), 494–500.

Nelson, S. B. (2002). Cortical microcircuits: Diverse or canonical? *Neuron*, **36**(1), 19–27.

Niell, C. M., & Stryker, M. P. (2010). Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, **65**(4), 472–479.

O'Leary, T., Williams, A. H., Franci, A., & Marder, E. (2014). Cell types, network homeostasis, and pathological compensation from a biologically plausible ion channel expression model. *Neuron*, **82**(4), 809–821.

Okabe, S., Kim, H.-D., Miwa, A., Kuriu, T., & Okado, H. (1999). Continual remodeling of postsynaptic density and its regulation by synaptic activity. *Nature Neuroscience*, **2**(9), 804–811.

Panas, D., Amin, H., Maccione, A., Muthmann, O., van Rossum, M., Berdondini, L., & Hennig, M. H. (2015). Sloppiness in spontaneously active neuronal networks. *Journal of Neuroscience*, **35**(22), 8480–8492.

Ponce-Alvarez, A., Mochol, G., Hermoso-Mendizabal, A., De la Rocha, J., & Deco, G. (2020). Cortical state transitions and stimulus response evolve along stiff and sloppy parameter dimensions, respectively. *eLife*, **9**, e53268.

Raman, D. V., & O'Leary, T. (2021). Optimal plasticity for memory maintenance during ongoing synaptic change. *eLife*, **10**, e62912.

Reichardt, W. (1987). Evaluation of optical motion information by movement detectors. *Journal of Comparative and Physiological Psychology*, **161**(4), 533–547.

Rubin, A., Geva, N., Sheintuch, L., & Ziv, Y. (2015). Hippocampal ensemble dynamics timestamp events in long-term memory. *eLife*, **4**, e12247.

Rule, M. E., Loback, A. R., Raman, D. V., Driscoll, L. N., Harvey, C. D., & O'Leary, T. (2020). Stable task information from an unstable neural population. *eLife*, **9**, e51121.

Rule, M. E., & O'Leary, T. (2022). Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proceedings of the National Academy of Sciences, USA*, **119**(7), e2106692119.

Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, **58**, 141–147.

Rule, M. E., Sorbaro, M., & Hennig, M. H. (2020). Optimal encoding in stochastic latent-variable models. *Entropy*, **22**(7), 714.

Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B., & Shanechi, M. M. (2020). Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience*, **24**, 140–149.

Scholl, C., Rule, M. E., & Hennig, M. H. (2021). The information theory of developmental pruning: Optimizing global network architectures using local synaptic rules. *PLoS Computational Biology*, **17**(10), e1009458.

Schoonover, C. E., Ohashi, S. N., Axel, R., & Fink, A. J. (2021). Representational drift in primary olfactory cortex. *Nature*, **594**(7864), 541–546.

Skaggs, W., Knierim, J., Kudrimoti, H., & McNaughton, B. (1994). A model of the neural basis of the rat's sense of direction. *Advances in Neural Information Processing Systems*. https://proceedings.neurips.cc/paper/1994/file/024d7f84fff11dd7e8d9c510137a2381-Paper.pdf

Sussillo, D., Jozefowicz, R., Abbott, L., & Pandarinath, C. (2016). Lfads-latent factor analysis via dynamical systems. *arXiv preprint arXiv:160806315*.

Sweeney, Y., & Clopath, C. (2020). Population coupling predicts the plasticity of stimulus responses in cortical circuits. *eLife*, **9**, e56053.

Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., & Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, **420**(6917), 788–794.

Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., & Sethna, J. P. (2015). Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *Journal of Chemical Physics*, **143**(1), 010901. https://doi.org/10.1063/1.4923066.

Urai, A. E., Doiron, B., Leifer, A. M., & Churchland, A. K. (2022). Large-scale neural recordings call for new insights to link brain and behavior. *Nature Neuroscience*, **25**(1), 11–19.

Wen, S., Yin, A., Furlanello, T., Perich, M., Miller, L., & Itti, L. (2021). Rapid adaptation of brain–computer interfaces to new neuronal ensembles or participants via generative modelling. *Nature Biomedical Engineering*, https://doi.org/10.1038/s41551-021-00811-z.

Wilmes, K. A., & Clopath, C. (2019). Inhibitory microcircuits for top-down plasticity of sensory representations. *Nature Communications*, **10**(1), 5055.

Wirtshafter, H. S., & Disterhoft, J. F. (2022). In vivo multi-day calcium imaging of CA1 hippocampus in freely moving rats reveals a high preponderance of place cells with consistent place fields. *Journal of Neuroscience*, **42**(22), 4538–4554.

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews. Neuroscience*, **16**(8), 487–497.

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., El Gamal, A., & Schnitzer, M. J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, **16**(3), 264.

Additional information

Competing interests

None.

Author contributions

Sole author.

Funding

No funding was received for this work.

Acknowledgements

I thank Angus Chadwick, Mark van Rossum and Michael Rule for their comments and discussion. Some graphics in the illustrations were sourced from https://scidraw.io/.

Keywords

neural circuits, neural manifolds, parameter spaces, sloppines

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Peer Review History