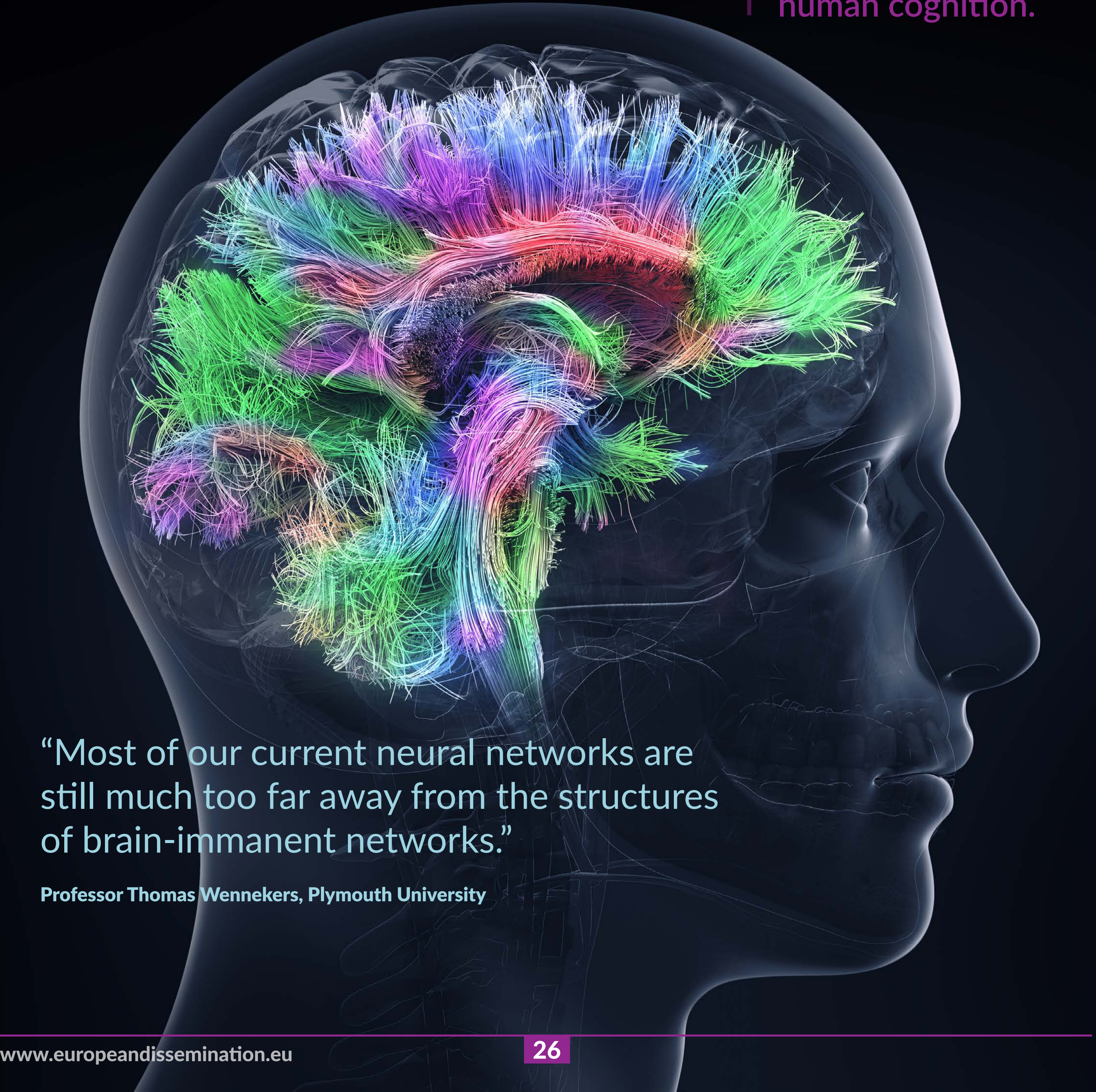
Material constraints enabling human cognition

Developing neural networks to unlock the secrets of human cognition.



"Most of our current neural networks are still much too far away from the structures of brain-immanent networks."

Professor Thomas Wennekers, Plymouth University

Why can we develop vocabularies consisting of tens of hundreds of thousands of words, yet our closest evolutionary relatives typically manage fewer than 100? This is just one of the vital, long-standing questions in cognitive science, linguistics and philosophy is set to tackle.

- How can humans build vocabularies of tens and hundreds of thousands of words, whereas our closest evolutionary relatives typically use fewer than 100?
- How is semantic meaning implemented for gestures and words, and, more specifically, for referential and categorical terms?
- How can grounding and interpretability of abstract symbols be anchored biologically?
- Which features of connectivity between nerve cells are crucial for the formation of discrete representations and categorical combination?
- Would modelling of cognitive functions using brain-constrained networks allow for better predictions on brain activity indexing the processing of signs and their meaning?

To find new answers to these questions, the MatCo project is utilising novel insights from human neurobiology and plans to translate these insights into mathematically exact computational models—neural network models.

The cognitive capacities of humans and higher mammals—their ability to learn, think, experience and sense—may depend on their brains' specific structural and functional features. If so, these neurobiological features must play a decisive role in explaining cognitive capacities.

Despite substantial progress in understanding brain function in general, explaining how structural and functional features of neural tissue bring about cognition, language and thought has remained a challenge.

Neural network models

Neural network models are potential tools for improving our understanding of complex brain functions.

A neural network is a network of interconnected neurone-like devices whose connections vary widely. Depending on the purpose of the simulation, they may be used to analyse a 'data set' using a process that imitates biological neurons signalling to each other, providing us with a simplified model of the human brain processing information.

To unlock the secrets of cognition, these models must be neurobiologically realistic. Despite neural networks advancing dramatically in recent years and even achieving human-like performance on complex perceptual and cognitive tasks, their similarity to aspects of brain anatomy and physiology is imperfect.

The MatCo team propose that neural networks for modelling cognition must incorporate a broad range of features that make them similar to real neurobiological networks at different levels: the microscopic level of nerve cell function, the mesoscopic level of interactions in local neuron clusters and the macroscopic level of interplay between these clusters and even larger brain parts and the whole brain.

Neural models of cognition explored

In their paper, 'Biological constraints on neural network models of cognitive function' (Pulvermüller *et al.*, 2021), featured in *Nature Reviews Neuroscience*, MatCo explore the different types of neural models of cognition and provide insight into how the biological plausibility of those models can be improved, i.e. how they can more closely mimic the functions within the human brain. Alongside the models themselves, MatCo has also identified a number of constraints that need to be applied to the models, as well as exciting future clinical applications of brain-constrained modelling.

Brain constraints

While increasing the neurobiological realism of the neural models is an important first step, a second crucial process is applying neuroscience constraints at different levels—the micro, meso- and macroscopic levels of description.

The novel proposed approach of 'brain-constrained' neural modelling aims at making 'neural' networks more neurobiologically plausible. The following seven subsections each deal with one specific aspect under which artificial neural models need to become more similar to real brains.

Integration at different levels

Previous modelling has mostly aimed to approximate neuronal function at the level of either single neurons (Gerstner and Naud, 2009; Teeter *et al.*, 2018), neuronal interaction in local cortical circuits (Schwalger, Deger and Gerstner, 2017; Malagarriga, Pons and Villa, 2019; Jansen and Rit, 1995; Potjans and Diesmann, 2014) or global interplay between cortical areas. To simultaneously apply constraints at different brain structure and function levels, these different levels must be addressed and integrated into a single model.

Neuron models

The functional units of the cortex and brain are neurons. All neural networks are composed of artificial correlates of neurons, but the level of detail with which neuronal function is simulated varies considerably (Gerstner and Naud, 2009; Teeter *et al.*, 2018; O'Reilly, Munakata and McClelland, 2000).

"Models that bridge the gap between the microscopic and macroscopic scales are a valuable resource in neuroscience."

Professor Friedemann Pulvermüller, Freie Universität Berlin.

The most detailed neuron model is not always the best choice for a given research question. While relatively basic neuron models yield excellent descriptions of neuronal activity (Gerstner and Naud, 2009), the greater computational resources required by sophisticated neuron models currently limit their applicability to large-scale simulations of within-area and across-area interactions relevant to cognition.

Synaptic plasticity and learning

The inclusion of learning mechanisms is a crucial ingredient of biologically plausible networks. However, localist and whole-brain models typically lack this feature. To model multiple learning systems in the brain, the implementation of both major forms of learning, supervised and unsupervised, is crucial.

Supervised learning presents a challenge—it requires feedback that informs the individual or network whether the performance was appropriate, wrong or erroneous. The choice of algorithms used in supervised learning simulations has been guided not only by biological plausibility (O'Reilly, 1998; Mollick, 2020) but also by the computational efficacy of gradient-dependent learning (Rumelhart, Hinton and Williams, 1986; Richards *et al.*, 2019; LeCun, Bengio and Hinton, 2015). Whether these latter algorithms are biologically realistic and applicable to sophisticated learning in specific cognitive areas is controversial.

Explicit feedback is important in some types of learning (such as reinforcement learning), and its biologically realistic implementation is crucial (O'Reilly, 1998; Cazin *et al.*, 2019; Mollick, 2020).

Inhibition and regulation

Brains are regulated systems. Cortical activity controls reasoning, emotion, thought, memory, language and consciousness and is regulated by control mechanisms at different levels. These include the microscopic local circuit level and the macroscopic, more global level of interacting brain parts, where cortical activity is regulated through information exchange with the thalamus, basal ganglia and other subcortical structures (Braitenberg, 1978; Yuille and Geiger, 2003; Gurney *et al.*, 2004). Many distributed neural networks simulating cognition are composed only of excitatory units, and they lack inhibition mechanisms (Schmidt *et al.*, 2018).

Inclusion of inhibition and regulation mechanisms at the local and more global levels is an important feature of making neurocognitive networks biologically plausible. Inhibitory neurons stop or restrain excitatory neurons from firing, providing gaps in activity. Without inhibition, the firing of neurons is ceaseless and disorganised. The rhythmic stop and start of electrical activity in the brain results in brain waves. Without a fine balance between this 'on and off' activity, brain waves become less coherent; a phenomena witnessed in psychiatric diseases, e.g. schizophrenia.

Area structure

The cortex is structured into a set of areas. Area definition is primarily based on anatomical criteria and sometimes refined using functional information. Depending on the question to be addressed by a simulation, a network model may implement one, a specific selection of or all cortical areas along with subcortical nuclei. Each area or nucleus can be realised as a separate 'layer' or model area, including a predefined number of artificial neurons. Dimensions of progressing towards biological realism include the range of brain parts and regions covered by the model. In the networks modelling language and conceptual processing, it is important to model a range of cortical areas known to be relevant for language and meaning.

Within-area local connectivity

Pyramidal cells are the most common excitatory neurons in the cortex. One of these cells may make contact with a few tens of thousands of other cortical cells within a pool of 15–32 billion neurons in the human cortex overall (Haug, 1987). Neuroanatomical studies indicate that local excitatory connections within a cortical area are sparse and show a neighbourhood bias towards links between adjacent neurons (Braitenberg and Schüz, 1998; Kaas, 1997).

Many networks that include auto-associative layers or areas (Willshaw, Buneman and Longuet-Higgins, 1969; Palm, 1982; Hinton and Shallice, 1991; Hopfield and Tank, 1985) include full connectivity between all neurons within these areas, which is not in line with the sparseness of intrinsic local cortical connections identified in the neuroanatomical studies. Hetero-associative networks lack the within-layer connections identified and, therefore do not seem biologically realistic either.

The brain constraint of sparse, local and partly random connections with a neighbourhood bias has been realised in some neural networks. Nonetheless, for most neural networks available today, the implementation of within-area connectivity constraints still leads to an increase in biological realism (van Albada *et al.*, 2020).

Between-area global connectivity

The connections between areas of the cortex follow some general rules. Most links are reciprocal. Adjacent areas are almost always interlinked, and second-next neighbours are connected in many cases (Braitenberg and Schüz, 1998; Young, Scannell, and Burns, 1995). However, longer-distance links are sparser, and much effort has been spent mapping them precisely using invasive and non-invasive techniques (van Albada *et al.*, 2020; Eichert *et al.*, 2019; Rojkova, 2016; Fernández-Miranda *et al.*, 2015, Rilling, 2014; Petrides *et al.*, 2012; (de Schotten *et al.*, 2012; Ardesch *et al.*, 2019; Barbeau, Descoteaux, and Petrides, 2020).

If two areas are interlinked, their connections are, in most cases, reciprocal and show topographic projections and local neighbouring relationships are preserved. Between-area connections are carried by long axon branches of cortical pyramidal cells. These axon branches pass through the white matter and can reach neurons in distant areas, where they branch and make contact with a local neighbourhood of neurons.

Essential brain constraints on artificial neural networks come from the connectivity structure of between-area links, as documented by neuroanatomical research.

Conclusion

The MatCo project targets novel biological explanations of specifically human cognitive and language abilities based on neurocomputational network simulations with networks similar to the structure and function of the relevant brain parts. Similarity between brains and networks needs to be constrained in at least seven ways, as discussed in the preceding subsections. By engineering cognitive mechanisms in a brain-constrained environment, the mechanisms underlying symbol learning, meaning acquisition, combinatorial learning and conceptual thought may become more graspable for human minds.

MatCo suggests a move towards more biologically oriented modelling, where neuroscience constraints have priority over other aims, such as processing efficacy and big data processing.

Models of brain function should attempt to integrate several and, ideally, all of the seven brain constraints explored by the team. The integration of microscopic and macroscopic levels is crucial to this endeavour.

Future focus

The work of the MatCo project offers very practical application strategies for the future. One such application addresses neuroplasticity, aiming at predicting and explaining the reorganisation of cognitive functions after a brain lesion or deprivation. With future potential for neurocomputational modelling constrained by specific features of an individual's brain, such insights could contribute to future planning of personalised therapy.

MatCo also continues to research and work in areas such as verbal working memory in the human brain, semantic binding between words and referent objects and actions, and the process of concrete and abstract concepts and meanings.

R References click here

PROJECT NAME

Material Constraints Enabling Human Cognition (MatCo)

PROJECT SUMMARY

Compared to our closest living relatives, who typically use fewer than 100 words, humans can build vocabularies of tens of hundreds of thousands of words. The ERC-funded Advanced Grant project 'Material Constraints enabling Human Cognition', or 'MatCo', will find out why. It will use novel insights from human neurobiology. These will be translated into mathematically exact computational models to find new answers to long-standing questions in cognitive science, linguistics and philosophy. The project will also explore how semantic meaning is implemented for gestures and words and, more specifically, for referential and categorical terms. To identify human cognitive capacities, MatCo will develop models replicating structural differences between human and non-human primate brains. The results will shed light on the biologically constrained networks.

PROJECT LEAD PROFILE

Friedemann Pulvermüller is Professor of Neuroscience of Language and Pragmatics at the Department of Philosophy of the Freie Universität Berlin, PI at the Berlin School of Mind and Brain at the Einstein Center of Neuroscience Berlin and at the Research Cluster 'Matters of Activity' of the German Research Foundation at the Humboldt University. He had taken PhDs in linguistics and psychology at the universities of Tübingen and Konstanz, before joining the Medical Research Council's Cognition and Brain Sciences Unit at Cambridge University as a Programme Leader in the Neuroscience of Language in 2000. In 2011, he moved to the Freie Universität to direct the Brain Language Laboratory Berlin. He has published over 300 publications, including a book on 'Neuroscience of Language' (Cambridge University Press, 2003).

PROJECT CONTACTS

Friedemann Pulvermüller

friedemann.pulvermuller@fu-berlin.de

www.fu-berlin.de/matco

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883811.

Material constraints enabling human cognition

van Albada, S. J., Morales-Gregorio, A., Dickscheid, T., Goulas, A., Bakker, R., Bludau, S., Palm, G., Hilgetag, C-C. and Diesmann, M. (2020) 'Bringing anatomical information into neuronal network models', arXiv:1312.6026.

Ardesch, D. J., Scholtens, L. H., Li, L., Preuss, T. M., Rilling, J. R. and van den Heuvel, M. P. (2019) 'Evolutionary expansion of connectivity between multimodal association areas in the human brain compared with chimpanzees', *Proceedings of the National Academy of Sciences USA*, 116, pp. 7101–7106. doi: 10.1073/pnas.1818512116.

Barbeau, E. B., Descoteaux, M. and Petrides, M. (2020) 'Dissociating the white matter tracts connecting the temporo-parietal cortical region with frontal cortex using diffusion tractography', *Scientific Reports*, 10, 8186. doi: 10.1038/s41598-020-64124-y.

Braitenberg, V. (1978) Cell Assemblies in the Cerebral Cortex, In: Heim, R. and Palm, G. (eds) *Theoretical Approaches to Complex Systems* (Lecture Notes in Biomathematics, vol. 21), Berlin: Springer, pp.171–188.

Braitenberg, V. and Schüz A. (1998) *Cortex: Statistics and Geometry of Neuronal Connectivity.* 2nd edn. Berlin: Springer.

Cazin, N., Alonso, M. L., Chiodi, P. S., Pelc, T., Harland, B., Weitzenfield, A., Fellous, J-M. and Dominey, P. F. (2019) 'Reservoir computing model of prefrontal cortex creates novel combinations of previous navigation sequences from hippocampal place-cell replay with spatial reward propagation', *PLOS Computational Biology*, 15, e1006624. doi: 10.1371/journal.pcbi.1006624.

Eichert, N. Verhagen, L., Folloni, D., Jbabdi, S., Khrapitchev, A. A., Sibson, N. R., Mantini, D., Sallet, J. and Mars, R. B. (2019) 'What is special about the human arcuate fasciculus? Lateralization, projections, and expansion', *Cortex*, 118, pp. 107–115. doi: 10.1016/j.cortex.2018.05.005.

Fernández-Miranda, J. C., Wang, Y., Pathak, S., Stefaneau, L., Verstynen, T. and Yeh, F. C. (2015) 'Asymmetry, connectivity, and segmentation of the arcuate fascicle in the human brain', *Brain Structure and Function*, 220(3), pp. 1665–1680. doi: 10.1007/s00429-014-0751-7.

Gerstner, W. and Naud, R. (2009) 'Neuroscience. How good are neuron models?', *Science*, 326, pp. 379-380. doi: 10.1126/science.1181936.

Gurney, K., Prescott, T. J., Wickens, J. R. and Redgrave, P. (2004) 'Computational models of the basal ganglia: from robots to membranes', *Trends in Neuroscience*, 27, pp. 453–459.

Haug, H. (1987) 'Brain sizes, surfaces, and neuronal sizes of the cortex cerebri: a stereological investigation of man and his variability and comparison with some mammals (primates, whales, marupials, insectivores, and one elephant)', *American Journal of Anatomy*, 180, pp. 126–142.

Hinton, G. E. and Shallice, T. (1991) 'Lesioning an attractor network: investigation of acquired dyslexia', *Psychological Review*, 98, pp. 74–95.

Hopfield, J. J. and Tank, D. W. (1985) '"Neural" computation of decisions in optimization problems', *Biological Cybernetics*, 52, pp. 141–152.

Jansen, B. H. and Rit, V. G. (1995) 'Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns', Biological Cybernetics, 73, pp. 357–366. doi: 10.1007/BF00199471.

Kaas, J. H. (1997) 'Topographic maps are fundamental to sensory processing', *Brain Research Bulletin*, 44, pp. 107–112.

LeCun, Y., Bengio, Y. and Hinton, G. (2015) 'Deep learning', *Nature*, 521, pp. 436–444.

Malagarriga, D., Pons, A. J. and Villa, A. E. (2019) 'Complex temporal patterns processing by a neural mass model of a cortical column', *Cognitive Neurodynamics*, 13, pp. 379–392. doi: 10.1007/s11571-019-09531-2.

Mollick, J. A., Hazy, T. E., Krueger, K. A., Nair, A., Mackie, P., Herd, S. A. and O'Reilly, R. C. (2020) 'A systems-neuroscience model of phasic dopamine', *Psychological Review*, 127(6), pp. 972–1021. doi: 10.1037/rev0000199.

O'Reilly, R. C. (1998) 'Six principles for biologically based computational models of cortical cognition', *Trends in Cognitive Science*, 2, pp. 455–562.

O'Reilly, R., Munakata, Y. and McClelland, J. (2000) *Computational explorations in cognitive neuroscience.* Cambridge, Mass.: MIT Press.

Palm, G. (1982) Neural assemblies. Berlin: Springer.

Petrides, M., Tomaiuolo, F., Yeterian, E. H. and Pandya, D. N. (2012) 'The prefrontal cortex: comparative architectonic organization in the human and the macaque monkey brains' *Cortex*, 48, pp. 46–57. doi: 10.1016/j.cortex.2011.07.002.

Potjans, T. C. and Diesmann, M. (2014) 'The cell-type specific cortical microcircuit: relating structure and activity in a full-scale spiking network model', *Cerebral Cortex*, 24, pp. 785–806. doi: 10.1093/cercor/bhs358.

Pulvermüller, F., Garagnani, M., and Wennekers, T. (2014) 'Thinking in circuits: toward neurobiological explanation in cognitive neuroscience', *Biological Cybernetics*, 108(5), 573-593. doi: 10.1007/s00422-014-0603-9.

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R. and Wennekers, T. (2021) 'Biological constraints on neural network models of cognitive function', *Nature Reviews Neuroscience*, 22, pp. 488–502. doi: 10.1038/s41583-021-00473-5.

Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., Clopath, C., Costa, R. P., de Berker, A., Ganguli, S., Gillon, C. J., Hafner, D., Kepecs, A., Kriegeskorte, N., Latham, P., Lindsay, G. W., Miller, K. D., Naud, R., Pack, C. C., Poirazi, P., Roelfsema, P., Sacramento, J., Saxe, A., Scellier, B., Schapiro, A. C., Senn, W., Wayne, G., Yamins, D., Zenke, F., Zylberberg, J., Therien, D. and Kording, K. P. (2019) 'A deep learning framework for neuroscience', *Nature Neuroscience*, 22, pp. 1761–1770. doi: 10.1038/s41593-019-0520-2.

Rilling, J. K. (2014) 'Comparative primate neuroimaging: insights into human brain evolution', *Trends in Cognitive Sciences*, 18, pp. 46–55.

Rojkova, K., Volle, E., Urbanski, M., Humbert, F., Dell'Acqua, F., and Thiebaut de Schotten, M. (2016) 'Atlasing the frontal lobe connections and their variability due to age and education: a spherical deconvolution tractography study', *Brain Structure and Function*, 221(3), pp. 1751–1766. doi: 10.1007/s00429-015-1001-3.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986) 'Learning Representations by Back- Propagating Errors', *Nature*, 323, pp. 533–536.

Schmidt, M., Bakker, R., Shen, K., Bezgin, G., Diesmann, M. and van Albada, S. J. (2018) 'A multi-scale layer-resolved spiking network model of resting-state dynamics in macaque visual cortical areas', *PLOS Computational Biology*, 14, e10006359.

Schomers, M.R., Garagnani, M. and Pulvermüller, F. (2017) 'Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex', *Journal of Neuroscience*, 37(11), 3045, doi: 10.1523/jneurosci.2693-16.2017.

de Schotten, M. T., Dell'Acqua, F., Valabregue, R. and Catani, M. (2012) 'Monkey to human comparative anatomy of the frontal lobe association tracts', *Cortex*, 48, pp. 82–96. doi: 10.1016/j.cortex.2011.10.001.

Schwalger, T., Deger, M. and Gerstner, W. (2017) 'Towards a theory of cortical columns: From spiking neurons to interacting neural populations of finite size', *PLOS Computational Biology*, 13, e1005507. doi: 10.1371/journal.pcbi.1005507.

Teeter, C., Iyer, R., Menon, V., Gouwens, N., Feng, D., Berg, J., Szafer, A., Cain, N., Zeng, H., Hawrylycz, M., Koch, C. and Mihalas, S. (2018) 'Generalized leaky integrate-and-fire models classify multiple neuron types', *Nature Communications*, 9, 709. doi: 10.1038/s41467-017-02717-4.

Tomasello, R., Garagnani, M., Wennekers, T. and Pulvermüller, F. (2017), 'Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex', *Neuropsychologia*, 98 (4), pp. 111–129. doi: 10.1016/j.neuropsychologia.2016.07.004

Tomasello, R., Wennekers, T., Garagnani, M. and Pulvermüller, F. (2019) 'Visual cortex recruitment during language processing in blind individuals is explained by Hebbian learning', *Scientific Reports*, 9(1), 3579. doi: 10.1038/s41598-019-39864-1.