# Making Sense of the Natural Environment

**Christoph von der Malsburg (FIAS, Frankfurt and INI, ETH Zürich), Benjamin Grewe (INI, ETH Zürich), Thilo Stadelmann (CAI, ZHAW Winterthur)**

The neural basis of cognition is unclear to this day. We here present a conceptual framework resolving the conflict Fodor & Pylyshyn (1988); Dever (2006) between symbolic and neural approaches. In our scheme, the cortical carriers of meaning are not individual neurons but sets of neurons supporting each other by mutual excitation. These sets and their supporting connectivity are called 'net fragments' or simply 'fragments.' Also fragments activate only as part of larger nets composed of overlapping fragments. Fragments play the role of composite symbols. As each neuron can be part of several fragments, and each fragment can overlap with several alternative other fragments, fragments can be likened to jigsaw puzzle pieces that fit together in innumerable different arrangements. Any such arrangement must, however, conform to a highly non-trivial consistency condition.

Net fragments and the composite nets they form are supported by specific patterns of synaptic connections. These are formed in development and learning by network self-organization, a process studied experimentally Goodhill (2007) and theoretically Willshaw & von der Malsburg (1979); Häussler & von der Malsburg (1983) on the example of the ontogenetic establishment of retinotopic fiber projections. This process selects net structures that are sparse (limited fan-in and fan-out of connections at each neuron) and are self-consistent such that a sufficient number of fibers converge on any one neuron from within the net. The composition rule for fragments to co-activate in a net is that together they form a net that is self-consistent (and would be stable under the process of network self-organization). Any particular large net (that is, set of active neurons) is unlikely to occur more than once in a life-time, so that only relatively small fragments have a chance to be active again and again to thus reach stability under network self-organization. But as these fragments overlap in multiple ways, cortex develops into an overlay of net fragments that supports an infinitude of consistent large-scale nets.

Among all possible thus-defined net structures a particular role is played by those that realize schema application. Each schema is an abstract structural description under which large numbers of instances can be united Bartlett (1932); Minsky (1974); Schank & Abelson (1977). Invariant object recognition has been modeled as schema application Arathorn (2002); Olshausen et al. (1995); Hinton (1981); Kree & Zippelius (1988); von der Malsburg (1988) realizable as a net that is representing schema, instance and the structure-preserving mapping between them Wolfrum et al. (2008). Natural intelligence may be defined as the ability of pursuing vital goals and intentions in varying contexts. Behavioral control has been classically described as schema application Shettleworth (2010). We propose the composition of nets out of fragments as basis for this process von der Malsburg et al. (2022).

In distinction to present-day artificial neural networks the human brain can learn and generalize from very few examples. It is a well-established insight Geman et al. (1992); Wolpert (1996) that such efficiency must be based on a deep structural relationship between learning system and domain. Inherent in our neural representation framework is therefore the claim that also the environment can be seen as a composite of a finite set of structural fragment types.

*Keywords*: neural representation, network self-organization, compositionality, net fragments, behavioral schema, intentions.

## References

Arathorn, D. (2002). *Map-seeking circuits in visual cognition – a computational mechanism for biological and machine vision*. Stanford, California: Standford Univ. Press.

Bartlett, F. (1932). *Remembering, a study in experimental and social psychology*. Cambridge: Cambridge University Press.

Dever, J. (2006). Compositionality. In E. Lepore & B. Smith (Eds.), *The oxford handbook of philosophy of language* (pp. 633–666). Oxford University Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, *28*(1), 3-71. doi: 10.1016/0010-0277(88)90031-5

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*, 1-58.

Goodhill, G. J. (2007). Contributions of theoetical modeling to the understanding of neural map development. *Neuron*, *56*, 301-311.

Häussler, A. F., & von der Malsburg, C. (1983). Development of retinotopic projections: An analytical treatment. *J. Theoretical Neurobiology*, *2*, 47–73. Retrieved from `https://vfs.fias.science/d/3cfce0fe5a/files/?p=/Retina.pdf`

Hinton, G. E. (1981). A Parallel Computation that Assigns Canonical Object-Based Frames of Reference. In *International joint conference on artificial intelligence* (pp. 683–685).

Kree, R., & Zippelius, A. (1988). Recognition of topological features of graphs and images in neural networks. *J. Phys. A*, *21*, 813-818.

Minsky, M. (1974, June). *A framework for representing knowledge* (Tech. Rep. No. 306). MIT AI Laboratory.

Olshausen, B., Anderson, C., & Van Essen, D. (1995). A multiscale dynamic routing circuit for forming size- and position-invariant object representations. *Journal of Computational Neuroscience*, *2*, 45-62.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. New Jersey: Erlbaum.

Shettleworth, S. (2010). *Cognition, evolution, and behavior (2nd ed.)*. Oxford: Oxford University Press.

von der Malsburg, C. (1988). Pattern recognition by labeled graph matching. *Neural Networks*, *1*, 141–148.

von der Malsburg, C., Stadelmann, T., & Grewe, B. (2022). *A theory of natural intelligence.* doi: 10.48550/ARXIV.2205.00002

Willshaw, D. J., & von der Malsburg, C. (1979). A marker induction mechanism for the establishment of ordered neural mappings; its application to the retinotectal problem. *Philosophical Transactions of the Royal Society of London, Series B*, *287*, 203–243.

Wolfrum, P., Wolff, C., Lücke, J., & von der Malsburg, C. (2008). A recurrent dynamic model for correspondence-based face recognition. *Journal of Vision*, *8*(7), 34. doi: 10.1167/8.7.34

Wolpert, D. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation*, *8*, 1341-1390.