



Toward understanding the neural code of the brain

Christoph von der Malsburg¹

Published online: 18 October 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

More and more, the neurosciences and the sciences concerned with mind and cognition are burying fundamental questions under layers of professional methodology. I therefore welcome Biological Cybernetics' invitation to comment on two of my papers, (von der Malsburg 1973) and (von der Malsburg and Schneider 1986) (henceforth referred to as (I) and (II)) as an opportunity to address two fundamental questions about brain and mind: How is the brain's structure generated? and How is mental content expressed by the brain's physical states? Those two questions are deeply entangled with each other and play a kind of gateway role on the way to making progress with the issues of perception, intelligence, creativity and consciousness.

Keywords Cognition · Symbols vs. Neurons · Binding issue · Network self-organization

Part I—How is the brain's structure generated?

Any fleeting thought in our mind is ultimately the result of the nested processes of evolution, ontogenesis, learning and education. Regarding all these and the whole context of understanding Life, a unifying theme and framework is just becoming perceptible as outline shining through a veil. For more than a hundred years, the discussion of evolution has been a battle between two attitudes, intelligent design and accumulation of “senseless” mutations. Neither of these helps to provide insight into Life's structure, at least as long as mutations are taken to be totally unstructured (Dawkins 1976) and the designing intelligence is seen as acting from outside the material world.

Since before the time of Darwin, biology in its various branches has brought to light highly systematic structure that characterizes living and past organisms and the rela-

tions between them and has, in recent decades, powerfully extended this perspective down to the molecular level, revealing systematic patterns of structure formation in cell and organism. Considering the layout of this landscape, the original front lines of the evolution discussion are no longer tenable. It is no longer too daring to say that Life is indeed the result of intelligent design, the design, however, happening under our eyes, as part of the game and within the material world.

To grasp its mechanisms, it may help to compare Life's structure to the architectures of technology that dominate the construction of buildings, of electronic chips, of software and of many more types of artifact. It is the essence of an architecture that the specification of concrete structures takes place in two stages, by first defining the architecture and then, within the bounds of that architecture, determining specific structures. An architecture is not a physical thing but rather an abstract conceptual framework that can be materialized in a variety of ways. A software architecture in the form of a programming language complete with development environment, for instance, constrains individual programs to what is functionally useful. Likewise, the bauplan of mammals, formulated as a parametrized system of ontogenetic development, constitutes an architecture that constrains individual species but still leaves tremendous room for variation.

An architecture comprises an array of sub-processes that link up with each other and with the environment in a complex network of nested loops of signals, all acting together in the sense of organismic coherence. Mutual consistency

Communicated by Benjamin Lindner.

To highlight the scientific impact of our Journal over the last decades, we asked authors of highly influential papers to reflect on the history of their study, the long-term effect it had, and future perspectives of their research. We trust the reader will enjoy these first-person accounts of the history of big ideas in Biological Cybernetics.

✉ Christoph von der Malsburg
malsburg@fias.uni-frankfurt.de

¹ Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany

of those loops is a powerful constraint. Mathematics takes this to the extreme, admitting only structures in which all loops of deduction are totally free of self-contradiction. This constraint admits, for instance, only five Platonic bodies or only four number fields—real, complex, quaternion and octonion. Also, technical or biological architectures are structured such as to support only contradiction-free and seamlessly fitting parts and levels. Like mathematics, the functional architecture of Life can be expected to form a Platonic realm. It is structured not by external intelligence but by the system-immanent constraint of consistency, which discerns between viable and non-viable structures. This, then, is the true meaning of intelligent design, that material structures and processes, once they come near to consistent architectural structure, are fully drawn into it as if by magical force, like a dangling chain into the form of a perfect catenoid or like a potatoid soap bubble into a perfect sphere.

Let us apply this perspective to the functioning brain. It is an amalgam of at least two architectures, those that David Hume (1740/1975) referred to when writing that *reason* was the slave of the *passions*. The “passions,” our behavioral architecture, ultimately goes back to our single-celled ancestors and implements the strategies and drives serving the basic goals of survival and procreation. Originally, this architecture was all expressed in terms of molecular signals and interactions, which on the way to multi-cellular organisms were transformed into the network of neural and humoral interactions pervading our body. Our behavioral architecture has absorbed, over the eons, all the experience evolution has collected with our near and distant ancestors inside their ecological environment, and it may well be called the architecture of our life.

I will here concentrate on the other architecture, Hume’s “reason,” the slave of the passions, and on the two fundamental questions of how it arises and how it expresses itself in physical terms. Our whole organism, including the brain, is ontogenetically constructed on the basis of one gigabyte of genetic information, 3.3 billion nuclear bases worth 2 bits each (Consortium 2001). To describe the wiring of the brain as a list of connections, on the other hand, takes a petabyte—a million gigabytes—of information (for a synapse to select one of the 10^{10} neurons of the brain it takes $\log_2(10^{10}) = 33$ bits; for the 10^{14} synapses of cortex it accordingly takes $3.3 \cdot 10^{15}$ bits).

How is this tremendous information gap to be bridged? Is it that John Locke (1690/1997) is right and the brain is a *tabula rasa* at birth, its wiring containing little information, empty space to be filled only after birth by masses of information flooding in through the senses?

For one thing, we are certainly not born as a *tabula rasa*, being endowed at birth with the behavioral architecture just mentioned and with the infrastructure to take in and administer experiences, as pointed out by Kant (1781/1999). And

then, the flood of information coming in through our senses is very limited. The environment in which children grow up could be simulated with the tools of modern virtual reality on the basis of a few gigabytes of information (see, for instance Crytek’s game *Robinson: The Journey*, URL: <https://www.crytek.com/games/robinson>), and through our lifetime we anyway absorb information only at the very modest rate of a couple of bits per second (Landauer 1986), running up to altogether a gigabit if we live long. The information gap disappears if the information content of a complex structure is measured not in terms of the number of bits needed to *describe* it but in terms of the number of bits of the shortest algorithm by which it can be *generated* (Li and Vitányi 2008). This directs our attention at this mechanism or mechanisms, this “Kolmogorov algorithm,” by which our body and brain are made.

The brain’s construction proceeds in stages. The first stage is part and extension of the construction kit of the whole body. As such it is based on a three-dimensional lay-out of marker molecules, “morphogens.” These are produced by the cells of the embryo and in turn control the step-wise specification of the cells’ genetic control hierarchy (Zeitlinger and Stark 2010). In this stage of development and in this fashion, the brain’s various regions (nuclei, cortical areas, etc.) are created together with morphogenetic markers and marker gradients in mutual interaction. In stage two, the lay-out of morphogens and their gradients guide the axons that are extended by neurons, creating a first sketch of the brain’s wiring diagram (Sperry 1951). It may be surmised that Hume’s “passions,” the definition of behavioral patterns, are a direct result of these two stages and are in this sense directly determined by genetic control.

This genetically specified sketch amounts to something like a hull constraining further refinement. In stage three, which begins late in embryonic development and continues throughout the life of the individual, this refinement of connectivity takes place in the form of pruning of connections and of moving axonal terminal branches over short distances, breaking some connections and making others. In these movements, the terminal branches and synapses are guided exclusively by the signals that are observable in their neighborhood and the signals arriving along the fibers from their own source neurons. These signals are created spontaneously by the neurons themselves (only later to be complemented by sensory input). Thus, connectivity and activity condition each other, connectivity shaping activity patterns, activity patterns acting back on connectivity. This interactive loop runs its course until it converges on network patterns that stabilize themselves. To this process of network self-organization, I will refer by the name connectivity dynamics.

This is the point at which (I), the first of my two articles comes in. However, before I come to that, let me draw atten-

tion to the analogy between developmental stage three and the previous two. Also those had as their basis a loop of interaction, cells putting out signals (in the form of morphogens), morphogens acting back on the cells' differentiation and signaling. In that case, the distribution of signals is shaped by three-dimensional space, which imposes its structure on the movement of cells and the diffusion of signals. This dance of interacting signals and cell differentiation is set up by evolution such as to guide the emergence of a self-consistent spatial lay-out of those two, cellular differentiation and molecular signals. While this process is beholden to and constrained by the rigid 'network' of spatial neighborhood relations, stage three, in which signals are transported along neural fibers, is free to escape from the dictate of three-dimensional topology to enter a totally new realm of forms, constrained only by the law of consistency between signals and connections.

Self-organization of orientation sensitive cells in the striate cortex

Now that the stage is set, let me discuss the first of my two publications, "Self-Organization of Orientation Sensitive Cells in the Striate Cortex," which appeared in *Biological Cybernetics* when it still was called *Kybernetik* (von der Malsburg 1973) and referred-to as (I).

According to the observations (Hubel and Wiesel 1977) to be modeled, neurons in the primary visual cortex of some animal species respond to a line or edge that appears in their receptive field in the retina when the orientation of that line or edge is near to an 'optimal angle' specific to each neuron, and this optimal angle changes continuously with the position of neurons within the cortical plane. The purpose of the model (I) was to let these two observations, oriented receptive fields and continuous distribution of orientations within the cortical plane, develop by connectivity dynamics starting from a plausible initial state. Connectivity among the neurons in the cortical plane was assumed in (I) to be fixed, in the form of short-range excitation and longer-range inhibition. Development concerned the reorganization of the receptive fields of the neurons under the influence of sensory input patterns in the form of oriented edges or lines of light. An initially random profile of the afferent connection strengths within the receptive fields then changed to a final state in the form of an oriented bar. This development of individual receptive fields was modeled in a style that later and in probabilistic formulation came to be called expectation maximization (Dempster et al. 1977): on the basis of the current values of the afferent connections neurons decide whether to fire, and the firing neurons then mold their receptive field profile to the current input pattern. They do this by strengthening their active afferent connections (in Hebbian fashion) at the expense of currently non-active connections (keeping the sum of all afferent connections constant). In this respect,

the model (I) was quickly followed by two others (Nass and Cooper 1975) and (Perez et al. 1975).

But the model (I) went further, by letting neurons exchange lateral excitation and inhibition within the cortical plane, so that a neuron could only fire together with cortical neighbors. In consequence, the model converged to a final state in which sets of neighboring neurons in cortex acquired receptive fields with similar orientation, those sets succeeding, over the course of development, in establishing themselves on the basis of mutual consistency between the fixed neighborhood interaction and plastic afferent organization in the participating neurons. These overlapping sets, or 'net fragments,' thus form a continuous map from stimulus orientation to cortical position.

The model is an early example of network self-organization. Under the influence of structured input statistics, it shows convergence to a state in which firing statistics as shaped by the structure of connections in turn supports that same connectivity structure. It is to be admitted, though, that the model as such is not a generic example of network self-organization being, first, dominated strongly by input statistics and, second, many of its connections being fixed. The model's initial connectivity state—neighborhood connections in a two-dimensional cortical sheet and localized afferent connections from the retina (via thalamus) to cortex—can be interpreted as the result of stage two development and early stage three.

After its publication, the model (I) turned out in one respect to be in conflict with experimental data: it was shown (Wiesel and Hubel 1974) that not only orientation specificity of cortical neurons but also the generation of a continuous map of orientation onto the cortical plane takes place before visual experience, whereas the model (I) required structured visual stimulation of the retina. Several attempts at modeling orientation specificity or even map formation in the absence of visual input were subsequently published in response to this observation. One invoked pattern formation within the receptive field of cortical neurons (Linsker 1986), another relied on a highly regular arrangement of afferent fibers (Miller 1992). These two attempts had the problem of being critically dependent on rather precise regularities and parameter constellations which in view of the experimentally observed scatter of afferent fibers are unrealistic. A third attempt (von der Malsburg and Cowan 1982) relied on a vestige of orientation specificity already in the retina and a molecular marker mechanism to induce a regular map of it on the cortex. My own last (and to me, I confess, most convincing) attempt (Grabska-Barwinska and von der Malsburg 2008) at explaining prenatal development of orientation specificity and orientation map continuity relies on experimentally observed spontaneous pre-natal activity waves in the retinae (Meister et al. 1991) and structure formation induced by them in the cortex. In recent decades, there has

been a scatter of occasional publications on the subject without much coherence and mutual reference between them.

In the mean time, another study had for some time aroused more interest, the generation of receptive fields in the form of oriented Gabor wavelets on the basis of the statistics of natural images (Olshausen and Field 1996). This study itself is now superseded in public interest by the generation of Gabor-like receptive fields inside deep learning systems (Krizhevsky et al. 2017). After all this, the functional significance of the regular arrangement of orientation specificity is not clear (Horton and Adams 2005), appearing only in some species anyway. So the topic is now a scientific backwater. The deeper significance of (I) and all experimental and theoretical work of which it was part lies more in its laying the foundation for understanding the general phenomenon of network self-organization. This came into its own with the age-old subject of the ontogenesis of retinotopic connections.

Retinotopy

Being easy to understand, convenient to study experimentally as well as theoretically, and being a dominant theme of the nervous system's wiring, retinotopy is the paradigm of network self-organization, the hydrogen atom of brain organization. Two types of theory have long dominated thinking about the ontogenesis of retinotopic connections: one based on marker molecules, the other on neurons' electrical activity. I myself was originally an ardent proponent of the electrical activity version, then had a revelation that made me convert to marker molecules, and finally it became clear on the basis of experiments that both play their role. The issue thus sits at the intersection of stage two and stage three.

Stage one and two create the setting: retina and tectum as two-dimensional sheets of neurons with short-range excitatory connections. Molecular gradients guide fibers that grow from the retina to the tectum and establish a first, diffuse mapping as initial state. In the electrical activity version of retinotopic map formation, neurons in the retina generate spontaneous spike activity. This activity is organized by lateral short-range excitatory connections into local swarms of simultaneously firing neurons. These activity patterns are conveyed by the projection fibers to the tectum, where short-range connectivity equally generates local activity swarms. These pairs of activity clouds, one in retina, one in tectum, act back on the retino-tectal connections by synaptic plasticity: fibers are strengthened or generated by simultaneous firing in their retinal source and their tectal target neurons. The growth of a fiber is, however, compensated by the reduction in the strength of other connections out of the fiber's source neuron and of other connections into its target neuron. Thus, the local retinal activity clouds and the tectal activity clouds they induce concentrate strength into the connections

between them. For the first simple model for this process, see (Willshaw and von der Malsburg 1976).

Network self-organization

This, then, is the nucleus, the central mechanism, of network self-organization, the Kolmogorov algorithm of the brain: existing connections generate and shape clouds of activity, and connections within the clouds are strengthened or generated at the expense of competitors, thus modifying the existing connectivity: activity is shaped by connectivity, and connectivity is shaped by activity. This process of connectivity dynamics continues through many cycles until it converges to an attractor state, a connectivity structure that stabilizes itself. It has an inherent tendency to create global order: connectivity structures that are coherent. In the retino-tectal example, 'global order' and 'coherence' refer to connectivity in the form of a fully continuous map from all of the retina to all of the tectum, which in this context is called systems matching (Gaze and Keating 1972). This tendency to global order results from the fact that overlapping clouds of activity conspire in strengthening the connections within their overlap. Attractor networks therefore have topological structure: connectivity supports sets of overlapping activity clouds (somewhat analogous to the open sets of topological spaces) and those clouds condense connections into their overlaps.

Global coherence is, however, not guaranteed, and the process of network self-organization can get caught in local optima. In the retinotopy case, such local optima would be mutually incoherent part-maps of retinal regions to tectal regions. The danger of ending up in local optima is to be avoided by starting the process with fairly large activity clouds supported by diffuse connectivity and letting activity clouds and connectivity contract gradually, thus progressing in coarse-to-fine fashion to find a detailed and globally coherent connectivity pattern.

Elucidation of the mechanism by which retinotopy is established in ontogenesis was a joint venture by several groups of experimental and theoretical neuroscientists. It resulted in a clear-cut conclusion, according to which the gradients of chemical morphogens of stage two establish boundary conditions for stage three, in which the growth behavior of neural fibers and the electrical (or molecular!), see (Willshaw and von der Malsburg 1979)) signals carried by them interact iteratively to create the final connectivity pattern. These conclusions were captured in extensive simulations (Willshaw and von der Malsburg 1979) and a compact mathematical description (Häussler and von der Malsburg 1983). Conclusion reached and fight over, the groups of scientists dispersed to work on other problems, leaving little by way of durable traces in textbooks and curricula. For the most recent reviews, see (Simpson and Goodhill 2011) (Kirkby

et al. 2013). The field of neurogenetic studies seems to have fallen back, as commented on by Hiesinger (2021), on considering only the processes of stage one and two. This is important work, as it has the potential of bringing us nearer to understanding the neural structures at the basis of our behavioral architecture, Hume’s “passions.”

Realizing, however, that the lion’s share of the brain’s structure is generated by connectivity dynamics, the nearly total neglect, if not ignorance, of this process not only in experimental neuroscience circles (where experimental accessibility is the main concern) but also in the computational neuroscience community is crippling progress with the problem of understanding the nature of human intelligence. See (von der Malsburg 2018) for a perspective on the great potential of network self-organization for understanding the brain’s function.

At least for me, the entry point to the subject of network self-organization was modelling the ontogenesis of orientation maps in visual cortex (I), and when David Willshaw came to the Max Planck Institute for biophysical Chemistry in Göttingen in 1973, he introduced me to the then highly contentious retinotopy topic, which we attacked together, first with the electrical activity version (Willshaw and von der Malsburg 1976), then with a molecular version thereof, the “marker induction theory” (von der Malsburg and Willshaw 1977), and finally with extensive simulations (Willshaw and von der Malsburg 1979) accounting for the full variety of seemingly contradictory experimental results available then. For me, the final point of that period of my scientific life was the development of a concise description of retinotopy development in the form of coupled differential equations for synaptic growth, complete with linear stability analysis and nonlinear coarse-to-fine coupling of the linear modes (Häussler and von der Malsburg 1983), something that would not have come about without Alexander Häussler, a Swiss mathematician who came as postdoc to my institute and who sadly passed away shortly after completing this work. The equations, deservedly called Häussler equations, have repeatedly proved useful for describing aspects of brain function, see for instance, (Zhu et al. 2010) or (Fernandes and von der Malsburg 2015).

Part II—How is mental content expressed by the brain’s physical states?

Understanding the brain’s function and the emulation thereof in *silico* are severely hampered by one as yet unresolved issue: how do the physical processes in the brain (and how could the electronic processes in the computer) act as language to represent mental content? Regarding this question, the community is deeply split between a symbolic and a sub-symbolic camp.

The symbolic camp is constituted by classical Artificial Intelligence and by cognitive science (especially linguistic theory), whereas the sub-symbolic camp (‘connectionism,’ ‘artificial neural networks’) speaks of neurons and their connections. The symbolic camp insists on a data format that is compositional so that a data item (such as the sentence ‘John loves Mary’) can exert effect on the basis of its structure and is systematic and productive in the sense of giving rise to analogous structures, (such as ‘Mary loves John’ or ‘Peter loves Edith’) (Fodor and Pylyshyn 1988).

This camp criticizes connectionism, whose data structure of pools of simultaneously active neurons offers on this front unsatisfactory choices: To represent ‘John loves Mary,’ it can either devote a neuron to the whole sentence, alternatively represent it as a ‘bag of features’—as the pool of neurons {‘John’, ‘loves’, ‘Mary’}—or try to represent the syntactical structure by neurons ‘John-subject,’ ‘loves-verb,’ ‘Mary-object.’ Neither of these is satisfactory. The bag-of-features version is ambiguous (being indistinguishable from ‘Mary loves John’), and the other two possibilities, based on combination-coding neurons, hamper structural generalization (e.g., from ‘John loves Mary’ to ‘Mary loves John’ and so on). The sub-symbolic neural camp retorts by criticizing the symbolic camp for being too intimately married to the sequential computer, for being too digital, for being unable to relate to the brain, and for being unable to account for learning.

The stand-off and lack of synthesis between the camps may be the main roadblock on the way to understanding brain function and true AI. This lack of progress may currently be hidden behind the general excitement over recent spectacular progress with problems like object classification from photographs (Krizhevsky et al. 2017) or speech-to-text conversion and natural language translation. This progress is based on a mixture of old (Rosenblatt 1961; Fukushima 1980; Rumelhart and McClelland 1986) and new (Hochreiter and Schmidhuber 1997; Vaswani et al. 2017) ideas, all of which are generally taken to be part of the neural camp. That camp can, however, not claim full victory, and the basic flaws of the sub-symbolic approach are coming back with a vengeance. Due to the underlying data structure’s inability to serve as basis for generalization—from an image of an object to transformed versions thereof, for instance, or from a sentence to another one describing the same actual situation—learning in current systems is extremely inefficient, needs huge masses of data and is restricted to interpolation between the samples already seen. For a recent discussion, see (Goyal and Bengio 2020). This is in sharp contrast to learning in humans, who are able to learn and generalize from a phenomenon after single exposure.

It seems highly desirable, then, to understand how the brain’s neural tissue and activity serves as data structure of the mind. This data structure evidently unites the strengths

of the symbolic and sub-symbolic versions discussed so far. The sub-symbolic camp's generally accepted version of this data structure, taking individual neurons as elementary symbols, undeniably is part of the truth, as it is resting on a solid experimental basis. But something seems to be lacking. A memorandum written four decades ago (von der Malsburg 1981) formulates this missing aspect as the "binding problem." The term refers to the putative mechanism that enables the brain to agglomerate neurons into a hierarchy of composite mental structures. The binding problem has gained public attention (Roskies 1999) but to this day no solution or even formulation has gained broad acceptance.

Sensory segmentation models

This is the context in which the second of my papers (von der Malsburg and Schneider 1986), referred-to here as (II), is to be discussed. It addresses a special case of the binding problem, the definition and representation of a compact sensory phenomenon and its distinction from background. In the visual modality, one speaks of the separation of figure from ground, whereas when the sensory phenomenon is a human voice one speaks of the cocktail-party problem (Cherry 1953). The problem is to both identify all sensory elements that belong to the phenomenon and to represent the result (the Gestalt in the parlance of the eponymous movement (Ellis 1950)) such that it can be treated as a composite whole.

The inner ear acts as a filter bank that decomposes the sound signal into frequency components. The human voice has harmonic structure and is composed of a series of components whose frequencies are multiples of the fundamental frequency. In (II), synthetic data are given in the form of an overlay of several harmonic spectra representing different simultaneously spoken human voices. The voices have different fundamental frequencies, and the amplitudes of all spectral components of one voice are modulated, sharing the same time course, as for instance the voice onset after explosiva. The model system is composed of two types of neurons, E-neurons, each of which is activated by one auditory frequency component and exerts excitation on other neurons, and a pool of H-neurons that are activated by the excitatory neurons and in turn inhibit them. The E-neurons respond to sensory input with sequences of bursts in relaxation oscillator fashion, whereas the H-pool passively follows its collective excitatory input.

The goal of the model is to bind together the E-neurons that belong to one voice by synchronizing their oscillations and desynchronizing them from all other E-neurons. The basis for this synchronization and desynchronization is rapid modification of the synapses between the E-neurons. If the activity in two neurons is positively correlated, their connection is strengthened; if it is negatively correlated, their connection

is weakened. This positive feedback loop between synchronization (desynchronization) of E-neuron oscillations and strengthening (weakening) of mutual coupling between them is set in motion by sensory signal correlations, especially common onset, between spectral components belonging to one voice. The strengthening of connections within sets of E-neurons that are positively correlated helps to couple those oscillators still further, thus leading to network self-organization.

As a result, the neurons representing the spectral components of one voice get coupled as a block and oscillate in perfect synchrony. This synchronization between all components of one voice can serve as basis for higher centers of the auditory system to focus attention on this voice, undisturbed by all other voices of the cocktail-party. All that is necessary, as briefly discussed in (II), is that all neurons involved in this analysis in higher centers pick up the rhythm of the sensory neurons, oscillate in sync with them, and with the help of the same kind of rapid synaptic plasticity just described strengthen their connections with the sensory neurons and decouple from the background neurons. The model has led in the hands of a student visitor, Avery Wang, to a technical application, the extraction of a single voice from background sound on the basis of harmonic structure and frequency modulation (Yue et al. 1998).

The model (II) is still far from establishing a general mechanism of sensory segmentation. It contents itself with a single type of evidence, temporal signal structure, and forgoes support by long-term learning and memory. A model of sensory segmentation in the olfactory cortex (Wang et al. 1990) goes to the other extreme, basing itself exclusively on memory. It assumes that individual odors have been recorded in the past by pair-coupling of all neurons responding to a single odor in the style of associative memory (Hopfield 1982). When a mixture of a small number of those individual odors is presented, oscillatory responses of the activated neurons are synchronized within known (recorded) odors and decorrelated between them, thus making it possible for the rest of the brain to focus attention on individual odor components by synchronizing with them.

The restriction in (II) to just one type of evidence was lifted in (Schneider 1986) (which has never been published in English and of which (II) actually represents only the first chapter). The same conceptual structure was applied in (von der Malsburg and Buhmann 1992) to the problem of visual segmentation. In both models, there is a field of cortical neurons with a tendency to oscillate. Each neuron relates to a one-dimensional (auditory case) or two-dimensional (visual case) position x and represents a specific feature type f . The system has a simple structure of excitatory connections, two neurons being connected if they agree in x or if they agree in f , and there are inhibitory pools that prevent global correlation. On the basis of feature distributions that are rather

homogeneous within segments and are different between them, both systems were able to subdivide simple synthetic input patterns into segments with correlated activity within and de-correlated activity between them.

Both models relate to coherent phenomena in the external world, human voices or solid objects, which create acoustic or optic disturbances that are analyzed by sensory organs, eye and ear, into fields of sensory signals and by neural filters into x -local feature types f . This analysis into components by our sensory apparatus poses the binding problem, the necessity to tie together and represent as one whole all those components that relate to the same external phenomenon, while at the same time keeping separate components that relate to different external phenomena. The neural connections of the system can be seen as modeling the physical interactions within the external phenomena, voice or solid object.

The binding issue

The report (von der Malsburg 1981) and the model (II) took a little while to attract attention, but then a response came in the form of first intriguing experimental hints (Eckhorn et al. 1988) and (Singer and Gray 1995) at the realization of binding by synchrony. Soon, however, the idea aroused violent opposition in the USA, possibly triggered by an editorial article in *Science Magazine* (Barinaga 1990). The most vociferous opponent to the idea, Anthony Movshon, organized a symposium during the 1993 Neuroscience Meeting of the Society for Neuroscience in Washington, DC, with the intent to critically discuss the idea of oscillations for feature binding, while others (Shadlen and Newsome 1998) created models to argue that cortical neurons could not process the temporal signal structure required for binding by synchrony. The discussion came to a head in a special issue of *Neuron* (Roskies 1999). In his contribution to that issue (Shadlen and Movshon 1999), Movshon proposed the challenge to experimentally demonstrate the object-global signal synchrony that seemed to be required by the theory of binding by synchrony. That experimental demonstration never materialized, discrediting in the eyes of many the idea of binding by synchrony.

Unfortunately, the heated debate about the subject let the community throw out the baby with the bath and lose interest in the binding problem altogether. It is true that the special problem of representing figure-ground separation can be solved within the framework of classical neural networks (Wersing et al. 2001), but the need for interpreting neural tissue as cognitive data structure goes way beyond figure-ground separation. The lack for a generally accepted solution to this problem may be the main roadblock on the way to understanding the mind and emulating it in the computer.

The article (II) must be seen against the background of the discussion (von der Malsburg 1981), which had raised the

binding issue and had proposed a solution in the dual form of signal synchrony and rapid synaptic plasticity. As a purely conceptual discussion, that report was rather inaccessible to the neuroscientific community. The model (II) was an attempt to present a simple and concrete application of its main ideas. Unfortunately, the impact of (II) turned out to be a mixed blessing. It illustrated the binding issue and its solution in the case of a simple example, the figure-ground problem, and it opened the way to experimental test, but in spite of first encouraging results the experiments eventually were judged by the community to be unconvincing. This had the negative effect that the binding issue was banned from the agenda of the neurophysiology community, leaving only a tradition of studying neural oscillations in various frequency bands. The binding issue as original motivation for this tradition has been suppressed mostly to the subconscious level, but see (von der Malsburg et al. 2010).

It is remarkable that the other half of the original proposal (von der Malsburg 1981), rapid changes in connectivity, quite central also in (II), got completely bypassed by the literature. The reason for this may be the experimental difficulty of measuring neural connection weights, let alone their rapid change in a situation-dependent way.

Part III—The emerging neural code

Another effect of emphasis on experimental accessibility was the concentration on very simple binding structures, figure-ground separation as well as feature binding (Treisman 1999). This in turn let the computational neuroscience community altogether miss the opportunity of understanding the binding issue as the missing link between neural and symbolic approaches to cognitive science and artificial intelligence.

This link is probably best established by considering schema application, a process that is central to cognition and that is best understood as a binding structure. In philosophy and psychology, there are proposals (Kant 1781/1999; Piaget 1923; Bartlett 1932) to understand structures and processes by reference to abstract schemata, in artificial intelligence by reference to ‘scripts’ or ‘frames’ (Schank and Abelson 1977; Minsky 1974). Abstract schemata are also the basis for analogy (Bartha 2019), case-based reasoning (Watson and Marir 1994), for the jurist’s interpretation of concrete cases relative to coded law or precedent or for the parsing of sentences by linguists. Schema application has even been proposed in an experimental context as basis for rapid learning (Tse et al. 2011).

The most concrete neural models for the process of schema application are dealing with visual object recognition based on template matching (von der Malsburg 1981; Bienenstock and von der Malsburg 1987; Lades et al. 1993; Olshausen

et al. 1993; Wiskott and von der Malsburg 1996; Arathorn 2002; Wolfrum et al. 2008). Here, a concrete object contained in an image is mapped onto an abstract template. Both the object-containing part of the image and the template are represented as two-dimensional fields of feature detector neurons, and the relation between them is established by a neural fiber mapping that preserves both neighborhood relations and feature type. Such mappings are called homeomorphic, a term borrowed from the mathematical field of topology. The template is invariant to changes in the position (and size and orientation) of the object in the image and may be abstract in more senses, relating for instance to only specific features in the image while ignoring others as irrelevant.

A specific neural model of schema-instance matching (Wiskott and von der Malsburg 1996) addresses the problem of invariant face recognition. A number of templates of faces of individual persons are represented as neural nets, a facial image of one of the individuals is presented in an image and the template of the correct person is to be activated. The difficulty of the problem lies in the fact that trial images can appear in any of an infinity of transformed versions—shifted or deformed (as by moderate depth rotation or change in expression)—so that rigid template matching is out of the question. In (Wiskott and von der Malsburg 1996), both face templates and input images are represented as two-dimensional fields of local feature detector neurons. The actual recognition is realized by a rapid process of network self-organization resulting in neighborhood-preserving one-to-one connections between corresponding points in image and template. The process starts from an initial state in which all neurons of the image field are connected to all neurons (of the same feature type) of the template field, irrespective of position. The feature-type specificity has the effect that points with similar texture in template and image are connected already in the initial state more densely than average.

The process of self-organization proceeds, in the model, exactly like in the retinotopy mechanism described above: Clouds of firing neurons are spontaneously formed in the image and template fields, and these clouds are local in those fields due to short-range excitation between neurons. Rapid synaptic plasticity strengthens synaptic connections between the two clouds and reduces the strengths of other connections running into or out of one of the clouds. Like in the retinotopy case, the process homes in to a one-to-one connectivity between the fields, neighboring neurons in the image field connecting to neighboring neurons in the template field. The active mappings thus developing can accommodate a fair amount of deformation. During the process, the different templates compete with each other, and the one with the best over-all point-to-point feature similarity with the image wins the process.

The model just described generates, and is based on, a binding structure that expresses both the neighborhood rela-

tions in object image and template and the point-to-point relations between object and template. This binding structure is generated by network self-organization and is expressed both in terms of connectivity (the fixed connections between neighboring neurons in image and template and the rapidly modifiable connections between image and template) and in terms of temporal signal correlations. The significance of this model is that it is binding structures of this kind that underlie the general process of schema-to-instance application.

It must be admitted that the system just described has a rather serious flaw—when taking into account realistic neural activation times (optimistically 5–10 msec for the formation of an activity cloud), it would take 10 or 100 seconds to recognize a face, orders of magnitude slower than in our brain. The system takes so much time because for each recognition attempt it has to run the full course of network self-organization from all-to-all (though feature specific) connectivity between the image and template fields to a one-to-one mapping.

The situation can, however, be changed drastically if traces of the connectivity structures generated by this slow process of network self-organization are permanently preserved and can be rapidly activated when needed, instead of generated *de novo*. This rapid activation of fiber projections could be neurally implemented with the help of control units (Anderson and van Essen 1987), which have been used in (Olshausen et al. 1993) for the purpose of object recognition. Whereas that model still needed engineered circuits to activate the control units, in the model (Wolfrum et al. 2008) control units were activated directly by the locally evaluated similarities between the activity pattern arriving on the set of fibers under the command of a single control unit and the activity pattern on the target neurons of those fibers. If in addition different control units cooperate appropriately through excitatory connections, whole coherent mappings can be activated in a single step. This is how the model (Wolfrum et al. 2008) solved the speed problem of the model (Wiskott and von der Malsburg 1996). The self-organization of the necessary connectivity has been demonstrated by simulation in the model (Fernandes and von der Malsburg 2015) and analyzed on the basis of the Häussler equation in (Zhu et al. 2010).

This face recognition model, or its technological implementation (Lades et al. 1993), is no longer state of the art (see, e.g., (Schroff et al. 2015)) and would have to be brought up to speed by further work. However, let us take it as a particular case and illustration of the general process of schema application, such as in parsing a sentence by relation to abstract syntactical structure, or performing an arithmetic calculation according to an algorithmic schema, or understanding a particular object as arrangement of parts, or analyzing a legal case in relation to precedent or code of law. The schema application is achieved by a dynamic process at the end of which stands a network of active connections homeomorphically

mapping the structured network representing the instance onto that representing the schema. Taking schema application as typical cognitive process (and assuming that the generalization from concrete models like (Wolfrum et al. 2008) is indeed viable), I would like to venture the claim that here lie the answers to the questions posed in the introduction: Mental content is represented by structured nets, and these nets are generated, in a nesting of time scales, as attractor states of connectivity dynamics. The thus conceived system is both neural, by being compatible with what we know about the brain, and symbolic, by being compositional, systematic and productive, as exacted by Fodor and Pylyshyn (1988).

In what way can one, from this perspective, still speak of binding? In the first of the two face recognition models (Wiskott and von der Malsburg 1996), binding was expressed by temporal signal correlations, but these served only to modify connectivity and at the end of the process reflected the connectivity structure. In the second model (Wolfrum et al. 2008), binding is exclusively expressed in terms of connectivity. As formulated there, the rapid connectivity dynamics underlying brain function no longer requires the time-consuming evaluation of signal correlations but consists in the activation of previously formed net fragments. Signal synchrony still plays a role, and a rather important one, as it supports life-long formation of new connectivity structure (such as when storing a new facial template).

Accepting the view outlined here (and explained in more detail in (von der Malsburg 2018)) implies a rather bold conclusion: All the symbolic structures ever generated in the brain are attractor states of connectivity dynamics. This may appear counter-intuitive, as it seems to impose a rather mechanical syntax on our inner world. The decisive growth condition (besides sparsity) for connectivity dynamics—mutual cooperation of alternative pathways between any two points in the nervous system, (and cooperation between pathways inside the nervous system and pathways in the external world)—goes, however, to the heart of what lets a mental state be viable: consistency of lines of reasoning (interpreting neural pathways as lines of reasoning). This condition of consistency of lines of reasoning is the strict criterion for a mathematical structure to be viable, and the example of mathematics demonstrates how potent that condition is in singling out structures that are interesting—and relevant to the world we live in (Wigner 1960)!

Conclusions

Setting out from two simple studies, of the generation of orientation domains in visual cortex (I) and of the cocktail-party effect (II), I have here spun out a discussion of two questions concerning brain and mind, the mechanisms by which our brain and our mind are made, and the interpretation of mate-

rial states of the nervous system as representation of mental structure. The current mainstream within the neuroscience and artificial intelligence communities gives rather unsatisfactory answers to these two questions, and it is likely that it is these unsatisfactory answers that are blocking progress toward understanding the function of the brain and emulating that function *in silico*.

As to the issue of representation, the criticism of Fodor and Pylyshyn (1988) raised against the neuroscientific view of mental representation (then going under the name of connectionism) still stands unanswered. Artificial Intelligence has, in transformer networks (Vaswani et al. 2017) evidently found a neural framework that is wide enough to represent mental content (interestingly containing an equivalent of dynamic connections, as pointed out, e.g., in (Goyal and Bengio 2020)), but that framework is totally passive in the sense of needing exhaustive training material and being unable to generalize beyond the examples it has seen.

This is in stark contrast to the ease with which children learn from comparatively minute amounts of data and rather restricted types of training material to then be able to generalize from this narrow basis and perceive, understand and behave in the world outside their nursery. This is only possible with the help of a very potent bias (Geman et al. 1992) which tunes the brain to the kind of world we live in (Wolpert 1996). Part of this bias is a behavioral repertoire (Eibl-Eibesfeldt 1996) that has been developed over the eons by evolution and is re-created in the individual under genetic control, presumably during what I have called stage two. Part of that bias is, however, of a more general nature and lifts our mind beyond the rather mechanical functions that, in most animals, are already apparent at birth. This more general bias I have discussed here as a result of connectivity dynamics and its tendency to create consistent webs of interdependent cognitive structures. Wigner marveled at the ‘unreasonable effectiveness’ of mathematics in describing the world (Wigner 1960), but the much greater marvel is the unreasonable effectiveness of the brain in perceiving the world (and, to boot, in discovering mathematical structures).

References

- Anderson C, van Essen D (1987) Shifter circuits: a computational strategy for dynamic aspects of visual processing. *PNAS* 84:6297–6301
- Arathorn D (2002) Map-Seeking circuits in Visual Cognition - A Computational Mechanism for Biological and Machine Vision. Stanford University Press, Stanford
- Barinaga M (1990) The mind revealed? *Science* 249:856–858. <https://doi.org/10.1126/science.2392677>
- Bartha P (2019) Analogy and analogical reasoning. In: Zalta EN (ed) *The stanford encyclopedia of philosophy*, Spring, 2019th edn. Stanford University, Metaphysics Research Lab
- Bartlett F (1932) *Remembering: A study in experimental and social psychology*. Cambridge University Press, Cambridge

- Bienenstock E, von der Malsburg C (1987) A neural network for invariant pattern recognition. *Europhys Lett* pp 121–126
- Cherry E (1953) Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* 25:975–979
- Consortium TIHG (2001) A physical map of the human genome. *Nature* 409:934–941. <https://doi.org/10.1038/35057157>
- Dawkins R (1976) *The Selfish Gene*. Oxford University Press, Oxford
- Dempster A, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *R Stat Soc B Met* 39(1):1–38
- Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck H (1988) Coherent oscillations: a mechanism for feature linking in the visual cortex? *Biol Cybern* 60:121–130
- Eibl-Eibesfeldt I (1996) Human ethology. In: Schmitt A, Atzwanger K, Grammer K, Schäfer K (eds) *New Aspects of Human Ethology*. Springer, Boston, MA, https://doi.org/10.1007/978-0-585-34289-4_1
- Ellis W (ed) (1950) *A source book of Gestalt psychology*. Routledge & Kegan Paul, London, <https://ia801600.us.archive.org/9/items/in.ernet.dli.2015.198039/2015.198039.A-Source-Book-Of-Gestalt-Psychology.pdf>
- Fernandes T, von der Malsburg C (2015) Self-organization of control circuits for invariant fiber projections. *Neural Comput* 27(5):1005–1032. https://doi.org/10.1162/NECO_a_00725
- Fodor JA, Pylyshyn ZW (1988) Connectionism and cognitive architecture: a critical analysis. *Cognition* 28(1):3–71. [https://doi.org/10.1016/0010-0277\(88\)90031-5](https://doi.org/10.1016/0010-0277(88)90031-5)
- Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by a shift in position. *Biol Cybern* 36:193–202. <https://doi.org/10.1007/BF00344251>
- Gaze R, Keating M (1972) The visual system and neuronal specificity. *Nature* 237:375–378. <https://doi.org/10.1038/237375a0>
- Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. *Neural Comput* 4:1–58
- Goyal A, Bengio Y (2020) Inductive biases for deep learning of higher-level cognition. *arXiv preprint arXiv:2011.15091*
- Grabska-Barwinska A, von der Malsburg C (2008) Perinatal ontogenesis of orientation specificity and maps in primary visual cortex of higher mammals. *J Neurosci* 28:249–257. <https://doi.org/10.1523/JNEUROSCI.5514-06.2008>
- Häussler A, von der Malsburg C (1983) Development of retinotopic projections—an analytical treatment. *J Theor Neurobiol* 2:47–73. <https://vfs.fias.science/d/3cfce0fe5a/files/?p=/Retina.pdf>
- Hiesinger P (2021) *The self-assembling brain: How neural networks grow smarter*. Princeton University Press
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hopfield J (1982) Neural networks and physical systems with emergent collective computational abilities. *PNAS* 79(8):2554–2558. <https://doi.org/10.1073/pnas.79.8.2554>
- Horton J, Adams D (2005) The cortical column: a structure without a function. *Philos Trans R Soc B* 360:837–862
- Hubel D, Wiesel T (1977) Ferrier-lecture: functional architecture of macaque monkey visual cortex. *Proc R Soc B* 198:1–59. <https://doi.org/10.1098/rspb.1977.0085>
- Hume D (1740/1975) *A treatise of human nature*. Clarendon Press, Oxford (Original work published in 1739–40)
- Kant I (1781/1999) *Critique of pure reason*. Cambridge University Press, Cambridge, England (Original work published in 1781)
- Kirkby L, Sack G, Firl A, Feller M (2013) A role for correlated spontaneous activity in the assembly of neural circuits. *Neuron* 80:1129–1144. <https://doi.org/10.1016/j.neuron.2013.10.030>
- Krizhevsky A, Sutskever I, Hinton G (2017) Imagenet classification with deep convolutional neural networks. *Commun ACM* 60:84–90. <https://doi.org/10.1145/3065386>
- Lades M, Vorbrüggen J, Buhmann J, Lange J, von der Malsburg C, Würtz R, Konen W (1993) Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans Comput* 42:300–311
- Landauer T (1986) How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Sci* 10(4):477–493. [https://doi.org/10.1016/S0364-0213\(86\)80014-3](https://doi.org/10.1016/S0364-0213(86)80014-3)
- Linsker R (1986) From basic network principles to neural architecture: emergence of orientation columns. *PNAS* 83:8779–8783
- Li M, Vitányi P (2008) Preliminaries. In: *An introduction to Kolmogorov complexity and its applications*, Springer, New York, pp 1–99. https://doi.org/10.1007/978-0-387-49820-1_1
- Locke J (1690/1997) *An essay concerning human understanding*. Penguin Classics, London (Original work published in 1690)
- Meister M, Wong R, Baylor D, Schatz C (1991) Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science* 252:939–943. <https://doi.org/10.1126/science.2035024>
- Miller KD (1992) Development of orientation columns via competition between ON- and OFF-center inputs. *NeuroReport* 3:73–76
- Minsky M (1974) *A framework for representing knowledge*. Memo 306, MIT-AI Laboratory
- Nass M, Cooper LN (1975) A theory for the development of feature detecting cells in visual cortex. *Biol Cybern* 19:1–18
- Olshausen B, Field D (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609
- Olshausen B, Anderson C, van Essen C (1993) A neurobiological model of visual attention and invariant recognition based on dynamic routing of information. *J Neurosci* 13:4700–4719
- Perez R, Glass L, Shlaer R (1975) Development of specificity in the cat visual cortex. *J Math Biol* 1:275–288
- Piaget J (1923) *Langage et pensée chez l'enfant*. Delachaux et Niestlé, Neuchâtel
- Rosenblatt F (1961) *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Spartan Books, Washington, D.C
- Roskies A (1999) The binding problem – review introduction. *Neuron* 24:7–9. [https://www.cell.com/neuron/issue?pii=S0896-6273\(00\)X0252-8](https://www.cell.com/neuron/issue?pii=S0896-6273(00)X0252-8)
- Rumelhart D, McClelland J (1986) *Parallel Distributed Processing*. MIT Press, Cambridge
- Schank R, Abelson R (1977) *Scripts, plans, goals and understanding: Inquiry into human knowledge structures*. Wiley, Nashville, TN
- Schneider W (1986) *Anwendung der Korrelationstheorie der Hirnfunktion auf das akustische Figur-Hintergrund-Problem (Cocktail-Party-Effekt)*. PhD thesis, Ruhr-Universität Bochum
- Schroff F, Kalenichenko D, Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), <https://doi.org/10.1109/CVPR.2015.7298682>
- Shadlen M, Movshon J (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24:67–77. [https://doi.org/10.1016/s0896-6273\(00\)80822-3](https://doi.org/10.1016/s0896-6273(00)80822-3)
- Shadlen M, Newsome W (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J Neurosci* pp 3870–96. <https://doi.org/10.1523/JNEUROSCI.18-10-03870.1998>
- Simpson HD, Goodhill G (2011) A simple model can unify a broad range of phenomena in retinotectal map development. *Biol Cybern* 104:9–29. <https://doi.org/10.1007/s00422-011-0417-y>
- Singer W, Gray C (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555–568

- Sperry R (1951) Regulative factors in the orderly growth of neural circuits. *Growth* 10:63–87
- Treisman A (1999) Solutions to the binding problem: progress through controversy and convergence. *Neuron* 24:105–10. [https://doi.org/10.1016/s0896-6273\(00\)80826-0](https://doi.org/10.1016/s0896-6273(00)80826-0)
- Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, Tohyama C, Bito H, Morris R (2011) Schema-dependent gene activation and memory encoding in neocortex. *Science* 333:891–895. <https://doi.org/10.1126/science.1205274>
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: *Advances in neural information processing systems*, pp 5998–6008
- von der Malsburg C (1973) Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* 14:85–100
- von der Malsburg C (2018) Concerning the neural code. *J Cog Sci* 19(4):511–550
- von der Malsburg C, Buhmann J (1992) Sensory segmentation with coupled neural oscillators. *Biol Cybern* 67:233–242
- von der Malsburg C, Cowan J (1982) Outline of a theory for the ontogenesis of iso-orientation domains in visual cortex. *Biol Cybern* 45:49–56
- von der Malsburg C, Schneider W (1986) A neural cocktail-party processor. *Biol Cybern* 54:29–40
- von der Malsburg C, Willshaw D (1977) How to label nerve cells so that they can interconnect in an ordered fashion. *PNAS* 74:5176–5178. <https://doi.org/10.1073/pnas.74.11.5176>
- von der Malsburg C (1981) The correlation theory of brain function. Internal report, 81-2, Max-Planck-Institut für Biophysikalische Chemie, Postfach 2841, 3400 Göttingen, FRG, reprinted in E. Domany, J.L. van Hemmen, and K.Schulten, editors, *Models of Neural Networks II*, chapter 2, pages 95–119. Springer, Berlin, 1994
- von der Malsburg C, Phillips W, Singer W (eds) (2010) *Dynamic coordination in the brain: Volume 5: From neurons to mind*. MIT Press, London, England
- Wang D, Buhmann J, von der Malsburg C (1990) Pattern segmentation in associative memory. *Neural Comput* 2:94–106
- Watson I, Marir F (1994) Case-based reasoning: a review. In: *The knowledge engineering review*, vol 9, Cambridge University Press, pp 327–354
- Wersing H, Steil JJ, Ritter H (2001) A competitive layer model for feature binding and sensory segmentation. *Neur Comput* 13(2):357–387
- Wiesel T, Hubel D (1974) Ordered arrangement of orientation columns in monkeys lacking visual experience. *J Comput Neurol* 158:307–318
- Wigner E (1960) The unreasonable effectiveness of mathematics in the natural sciences. *Communications in Pure and Applied Mathematics* 13 No. 1:1–14. <https://doi.org/10.1002/cpa.3160130102>
- Willshaw DJ, von der Malsburg C (1976) How patterned neural connections can be set up by self-organization. *Proc R Soc B* 194:431–445
- Willshaw DJ, von der Malsburg C (1979) A marker induction mechanism for the establishment of ordered neural mappings; its application to the retinotectal problem. *Philos Trans R Soc B* 287:203–243. <https://doi.org/10.1098/rstb.1979.0056>
- Wiskott L, von der Malsburg C (1996) Face recognition by dynamic link matching. In: Sirosh J, Miiikkulainen R, Choe Y (eds) *Lateral Interactions in the Cortex: Structure and Function*. Electronic book, The UTCS Neural Networks Research Group, Austin, TX, chap 11, <http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96/>
- Wolfrum P, Wolff C, Lücke J, von der Malsburg C (2008) A recurrent dynamic model for correspondence-based face recognition. *J Vision* 8(7):34. <https://doi.org/10.1167/8.7.34>
- Wolpert D (1996) The lack of a priori distinctions between learning algorithms. *Neural Comput* 8:1341–1390
- Yue LL, Dalal PB, Wang A (1998) Harmonic and frequency-locked loop pitch tracker and sound separation system. <https://patents.justia.com/patent/5812737>
- Zeitlinger J, Stark A (2010) Developmental gene regulation in the era of genomics. *Dev Biol* 339(2):230–239. <https://doi.org/10.1016/j.ydbio.2009.12.039>
- Zhu J, Bergmann U, von der Malsburg C (2010) Self-organization of steerable topographic mappings as basis for translation invariance. In: Diamantaras K, Duch W, Iliadis LS (eds) *Artificial neural networks – ICANN 2010*, Springer, Berlin, pp 414–419

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.