Differences Between Natural and Artificial Cognitive Systems

Wolf Singer

Contents

Introduction	17
Strategies for the Encoding of Relations: A Comparison Between Artificial and Natural Systems ..	19
Encoding of Relations in Feed-Forward Architectures	19
Encoding of Relations by Assemblies	20
A Comparison Between the Two Strategies	21
Assembly Coding and the Binding Problem	22
Computing in High-Dimensional State Space	22
Information Processing in Natural Recurrent Networks	24
Concluding Remarks	25
References	26

Abstract

This chapter identifies the differences between natural and artificial cognitive systems. Benchmarking robots against brains may suggest that organisms and robots both need to possess an internal model of the restricted environment in which they act and both need to adjust their actions to the conditions of the respective environment in order to accomplish their tasks. However, computational strategies to cope with these challenges are different for natural and artificial systems. Many of the specific human qualities cannot be deduced from the neuronal functions of individual brains alone but owe their existence to cultural evolution. Social interactions between agents endowed with the cognitive abilities of humans generate immaterial realities, addressed as social or cultural realities. Inten-

tionality, morality, responsibility and certain aspects of consciousness such as the qualia of subjective experience belong to the immaterial dimension of social realities. It is premature to enter discussions as to whether artificial systems can acquire functions that we consider as intentional and conscious or whether artificial agents can be considered as moral agents with responsibility for their actions.

Keywords

Artificial intelligence · Cognitive systems · Brain · Robot · Cultural evolution · Neuronal functions · Consciousness

W. Singer (✉)

Max Planck Institute for Brain Research (MPI), Ernst Strüngmann Institute for Neuroscience (ESI) in Cooperation with Max Planck Society, Frankfurt, Germany

Frankfurt Institute for Advanced Studies (FIAS), Frankfurt, Germany
e-mail: wolf.singer@brain.mpg.de

Introduction

Organisms and robots have to cope with very similar challenges. Both need to possess an internal model of the restricted environment in which they act and both need to

adjust their actions to the idiosyncratic conditions of the respective environment in order to accomplish particular tasks. However, the computational strategies to cope with these challenges exhibit marked differences between natural and artificial systems.

In natural systems the model of the world is to a large extent inherited, i.e. the relevant information has been acquired by selection and adaptation during evolution, is stored in the genes and expressed in the functional anatomy of the organism and the architecture of its nervous systems. This inborn model is subsequently complemented and refined during ontogeny by experience and practice. The same holds true for the specification of the tasks that the organism needs to accomplish and for the programs that control the execution of actions. Here, too, the necessary information is provided in part by evolution and in part by lifelong learning. In order to be able to evolve in an ever-changing environment, organisms have evolved cognitive systems that allow them to analyse the actual conditions of their embedding environment, to match them with the internal model, update the model, derive predictions and adapt future actions to the actual requirements.

In order to complement the inborn model of the world organisms rely on two different learning strategies: Unsupervised and supervised learning. The former serves to capture frequently occurring statistical contingencies in the environment and to adapt processing architectures to the efficient analysis of these contingencies. Babies apply this strategy for the acquisition of the basic building blocks of language. The unsupervised learning process is implemented by adaptive connections that change their gain (efficiency) as a function of the activity of the connected partners. If in a network two interconnected neurons are frequently coactivated, because the features to which they respond are often present simultaneously, the connections among these two neurons become more efficient. The neurons representing these correlated features become associated with one another. Thus, statistical contingencies between features get represented by the strength of neuronal interactions. “Neurons wire together if they fire together”. Conversely, connections among neurons weaken, if these are rarely active together, i.e. if their activity is uncorrelated. By contrast, supervised learning strategies are applied when the outcome of a cognitive or executive process needs to be evaluated. An example is the generation of categories. If the system were to learn that dogs, sharks and eagles belong to the category of animals it needs to be told that such a category exists and during the learning process it needs to receive feedback on the correctness of the various classification attempts. In case of supervised learning the decision as to whether a particular activity pattern induces a change in coupling is made dependent not only on the

local activity of the coupled neurons but on additional gating signals that have a “now print” function. Only if these signals are available in addition can local activity lead to synaptic changes. These gating signals are generated by a few specialized centres in the depth of the brain and conveyed through widely branching nerve fibres to the whole forebrain. The activity of these value assigning systems is in turn controlled by widely distributed brain structures that evaluate the behavioural validity of ongoing or very recently accomplished cognitive or executive processes. In case the outcome is positive, the network connections whose activity contributed to this outcome get strengthened and if the outcome is negative they get weakened. This retrospective adjustment of synaptic modifications is possible, because activity patterns that potentially could change a connection leave a molecular trace at the respective synaptic contacts that outlasts the activity itself. If the “now print” signal of the gating systems arrives while this trace is still present, the tagged synapse will undergo a lasting change (Redondo and Morris 2011; Frey and Morris 1997). In this way, the network’s specific activity pattern that led to the desired outcome will be reinforced. Therefore, this form of supervised learning is also addressed as reinforcement learning.

Comparing these basic features of natural systems with the organization of artificial “intelligent” systems already reveals a number of important differences.

Artificial systems have no evolutionary history but are the result of a purposeful design, just as any other tool humans have designed to fulfil special functions. Hence, their internal model is installed by engineers and adapted to the specific conditions in which the machine is expected to operate. The same applies for the programs that translate signals from the robot’s sensors into action. Control theory is applied to assure effective coordination of the actuators. Although I am not a specialist in robotics I assume that the large majority of useful robots is hard wired in this way and lacks most of the generative, creative and self-organizing capacities of natural agents.

However, there is a new generation of robots with enhanced autonomy that capitalize on the recent progress in machine learning. Because of the astounding performance of these robots, autonomous cars are one example, and because of the demonstration that machines outperform humans in games such as Go and chess, it is necessary to examine in greater depth to which extent the computational principles realized in these machines resemble those of natural systems.

Over the last decades the field of artificial intelligence has been revolutionized by the implementation of computational strategies based on artificial neuronal networks. In the second half of the last century evidence accumulated that relatively simple neuronal networks, known as Perceptrons or Hopfield

nets, can be trained to recognize and classify patterns and this fuelled intensive research in the domain of artificial intelligence. The growing availability of massive computing power and the design of ingenious training algorithms provided compelling evidence that this computational strategy is scalable. The early systems consisted of just three layers and a few dozens of neuron like nodes. The systems that have recently attracted considerable attention because they outperform professional Go players, recognize and classify correctly huge numbers of objects, transform verbal commands into actions and steer cars, are all designed according to the same principles as the initial three-layered networks. However, the systems now comprise more than hundred layers and millions of nodes which has earned them the designation “deep learning networks”. Although the training of these networks requires millions of training trials with a very large number of samples, their amazing performance is often taken as evidence that they function according to the same principles as natural brains. However, as detailed in the following paragraph, a closer look at the organization of artificial and natural systems reveals that this is only true for a few aspects.

Strategies for the Encoding of Relations: A Comparison Between Artificial and Natural Systems

The world, animate and inanimate, is composed of a relatively small repertoire of elementary components that are combined at different scales and in ever different constellations to bring forth the virtually infinite diversity of objects. This is at least how the world appears to us. Whether we are caught in an epistemic circle and perceive the world as composite because our cognitive systems are tuned to divide wholes into parts or because the world is composite and our cognitive systems have adapted to this fact will not be discussed further. What matters is that the complexity of descriptions can be reduced by representing the components and their relations rather than the plethora of objects that result from different constellations of components. It is probably for this reason that evolution has optimized cognitive systems to exploit the power of combinatorial codes. A limited number of elementary features is extracted from the sensory environment and represented by the responses of feature selective neurons. Subsequently different but complementary strategies are applied to evaluate the relations between these features and to generate minimally overlapping representations of particular feature constellations for classification. In a sense this is the same strategy as utilized by human languages. In the Latin alphabet, 28 symbols suffice to compose the world literature.

Encoding of Relations in Feed-Forward Architectures

One strategy for the analysis and encoding of relations is based on convergent feed-forward circuits. This strategy is ubiquitous in natural systems. Nodes (neurons) of the input layer are tuned to respond to particular features of input patterns and their output connections are made to converge on nodes of the next higher layer. By adjusting the gain of these converging connections and the threshold of the target node it is assured that the latter responds preferentially to only a particular conjunction of features in the input pattern (Hubel and Wiesel 1968; Barlow 1972). In this way consistent relations among components become represented by the activity of conjunction-specific nodes (see Fig. 1). By iterating this strategy across multiple layers in hierarchically structured feed-forward architectures complex relational constructs (cognitive objects) can be represented by conjunction-specific nodes of higher order. This basic strategy for the encoding of relations has been realized independently several times during evolution in the nervous systems of different phyla (molluscs, insects, vertebrates) and reached the highest degree of sophistication in the hierarchical arrangement of processing levels in the cerebral cortex of mammals (Felleman and van Essen 1991; Glasser et al. 2016; Gross et al. 1972; Tsao et al. 2006; Hirabayashi et al. 2013; Quiñero et al. 2005). This strategy is also the hallmark of the numerous versions of artificial neuronal networks designed for the recognition and classification of patterns (Rosenblatt 1958; Hopfield 1987; DiCarlo and Cox 2007; LeCun et al. 2015). As mentioned above, the highly successful recent developments in the field of artificial intelligence, addressed as “deep learning networks” (LeCun et al.

Fig. 1 The encoding of relations by conjunction-specific neurons (red) in a three-layered neuronal network. A and B refer to neurons at the input layer whose responses represent the presence of features A and B. Arrows indicate the flow of activity and their thickness the efficiency of the respective connections. The threshold of the conjunction-specific neuron is adjusted so that it responds only when A and B are active simultaneously

Fig. 2 Topology of a deep learning network

2015; Silver et al. 2017, 2018), capitalize on the scaling of this principle in large multilayer architectures (see Fig. 2).

Encoding of Relations by Assemblies

In natural systems, a second strategy for the encoding of relations is implemented that differs in important aspects from the formation of individual, conjunction-specific neurons (nodes) and requires a very different architecture of connections. In this case, relations among components are encoded by the temporary association of neurons (nodes) representing individual components into cooperating assemblies that respond collectively to particular constellations of related features. In contrast to the formation of conjunction-specific neurons by convergence of feed-forward connections, this second strategy requires recurrent (reciprocal) connections between the nodes of the same layer as well as feed-back connections from higher to lower levels of the processing hierarchy. In natural systems, these recurrent connections outnumber by far the feed-forward connections. As proposed by Donald Hebb as early as 1949, components (features) of

composite objects can not only be related to one another by the formation of conjunction-specific cells but also by the formation of functionally coherent assemblies of neurons. In this case, the neurons that encode the features that need to be bound together become associated into an assembly. Such assemblies, so the original assumption, are distinguished as a coherent whole that represents a particular constellation of components (features) because of the jointly enhanced activation of the neurons constituting the assembly. The joint enhancement of the neurons' activity is assumed to be caused by cooperative interactions that are mediated by the reciprocal connections between the nodes of the network. These connections are endowed with correlation-dependent synaptic plasticity mechanisms (Hebbian synapses, see below) and strengthen when the interconnected nodes are frequently co-activated. Thus, nodes that are often co-activated because the features to which they respond do often co-occur in the environment enhance their mutual interactions. As a result of these cooperative interactions, the vigour and/or coherence of the responses of the respective nodes is enhanced when they are activated by the respective feature constellation. In this way, consistent relations among the components of cognitive

objects are translated into the weight distributions of the reciprocal connections between network nodes and become represented by the joint responses of a cooperating assembly of neurons. Accordingly, the information about the presence of a particular constellation of features is not represented by the activity of a single conjunction-specific neuron but by the amplified or more coherent or reverberating responses of a distributed assembly of neurons.

A Comparison Between the Two Strategies

Both relation-encoding strategies have advantages and disadvantages and evolution has apparently opted for a combination of the two. Feed-forward architectures are well suited to evaluate relations between simultaneously present features, raise no stability problems and allow for fast processing. However, encoding relations exclusively with conjunction-specific neurons is exceedingly expensive in terms of hardware requirements. Because specific constellations of (components) features have to be represented explicitly by conjunction-specific neurons via the convergence of the respective feed-forward connections and because the dynamic range of the nodes is limited, an astronomically large number of nodes and processing levels would be required to cope with the virtually infinite number of possible relations among the components (features) characterizing real-world objects, leave alone the representation of nested relations required to capture complex scenes. This problem is addressed as the “combinatorial explosion”. Consequently, biological systems relying exclusively on feed-forward architectures are rare and can afford representation of only a limited number of behaviourally relevant relational constructs. Another serious disadvantage of networks consisting exclusively of feed-forward connections is that they have difficulties to encode relations among temporally segregated events (temporal relations) because they lack memory functions.

By contrast, assemblies of recurrently coupled, mutually interacting nodes (neurons) can cope very well with the encoding of temporal relations (sequences) because such networks exhibit fading memory due to reverberation and can integrate temporally segregated information. Assembly codes are also much less costly in terms of hardware requirements, because individual feature specific nodes can be recombined in flexible combinations into a virtually infinite number of different assemblies, each representing a different cognitive content, just as the letters of the alphabet can be combined into syllables, words, sentences and complex descriptions (combinatorial code). In addition, coding space is dramatically widened because information about the statistical contingencies of features can be encoded not only in the synaptic weights of feed forward connections but also

in the weights of the recurrent and feed-back connections. Finally, the encoding of entirely new or the completion of incomplete relational constructs (associativity) is facilitated by the cooperativity inherent in recurrently coupled networks that allows for pattern completion and the generation of novel associations (generative creativity).

However, assembly coding and the required recurrent networks cannot easily be implemented in artificial systems for a number of reasons. First and above all it is extremely cumbersome to simulate the simultaneous reciprocal interactions between large numbers of interconnected nodes with conventional digital computers that can perform only sequential operations. Second, recurrent networks exhibit highly non-linear dynamics that are difficult to control. They can fall dead if global excitation drops below a critical level and they can engage in runaway dynamics and become epileptic if a critical level of excitation is reached. Theoretical analysis shows that such networks perform efficiently only if they operate in a dynamic regime close to criticality. Nature takes care of this problem with a number of self-regulating mechanisms involving normalization of synaptic strength (Turri-giano and Nelson 2004), inhibitory interactions (E/I balance) (Yizhar et al. 2011) and control of global excitability by modulatory systems, that keep the network within a narrow working range just below criticality (Plenz and Thiagarajan 2007; Hahn et al. 2010).

The third problem for the technical implementation of biological principles is the lack of hardware solutions for Hebbian synapses that adjust their gain as a function of the correlation between the activity of interconnected nodes. Most artificial systems rely on some sort of supervised learning in which temporal relations play only a minor role if at all. In these systems the gain of the feed-forward connections is iteratively adjusted until the activity patterns at the output layer represent particular input patterns with minimal overlap. To this end very large samples of input patterns are generated, deviations of the output patterns from the desired result are monitored as “errors” and backpropagated through the network in order to change the gain of those connections that contributed most to the error. In multilayer networks this is an extremely challenging procedure and the breakthroughs of recent developments in deep learning networks were due mainly to the design of efficient backpropagation algorithms. However, these are biologically implausible. In natural systems, the learning mechanisms exploit the fundamental role of consistent temporal relations for the definition of semantic relations. Simultaneously occurring events usually have a common cause or are interdependent because of interactions. If one event consistently precedes the other, the first is likely the cause of the latter, and if there are no temporal correlations between the events, they are most likely unrelated. Likewise, components (features) that often occur together are likely to be related, e.g. because their particular constellation

is characteristic for a particular object or because they are part of a stereotyped sequence of events. Accordingly, the molecular mechanisms developed by evolution for the establishment of associations are exquisitely sensitive to temporal relations between the activity patterns of interconnected nodes. The crucial variable that determines the occurrence and polarity of gain changes of the connections is the *temporal relation* between discharges in converging presynaptic inputs and/or between the discharges of presynaptic afferents and the activity of the postsynaptic neuron. In natural systems most excitatory connections—feed forward, feed-back and recurrent—as well as the connections between excitatory and inhibitory neurons are adaptive and can change their gain as a function of the correlation between pre- and postsynaptic activity. The molecular mechanisms that translate electrical activity in lasting changes of synaptic gain evaluate correlation patterns with a precision in the range of tens of milliseconds and support both the experience-dependent generation of conjunction-specific neurons in feed-forward architectures and the formation of assemblies.

Assembly Coding and the Binding Problem

Although the backpropagation algorithm mimics in a rather efficient way the effects of reinforcement learning in deep learning networks it cannot be applied for the training of recurrent networks because it lacks sensitivity to temporal relations. However, there are efforts to design training algorithms applicable to recurrent networks and the results are promising (Bellec et al. 2019).

Another and particularly challenging problem associated with assembly coding is the *binding problem*. This problem arises whenever more than one object is present and when these objects and the relations among them need to be encoded within the same network layer. If assemblies were solely distinguished by enhanced activity of the constituting neurons, as proposed by Hebb (1949), it becomes difficult to distinguish which of the neurons with enhanced activity actually belong to which assembly, in particular, if objects share some common features and overlap in space. This condition is known as the *superposition catastrophe*. It has been proposed that this problem can be solved by multiplexing, i.e. by segregating the various assemblies in time (Milner 1992; von der Malsburg and Buhmann 1992; for reviews see Singer and Gray 1995; Singer 1999). Following the discovery that neurons in the cerebral cortex can synchronize their discharges with a precision in the millisecond range when they engage in high frequency oscillations (Gray et al. 1989), it has been proposed that the neurons temporarily bound into assemblies are distinguished not only by an increase of their discharge rate but also by the precise synchronization of their action potentials (Singer and Gray 1995; Singer 1999).

Synchronization is as effective in enhancing the efficiency of neuronal responses in down-stream targets as is enhancing discharge rate (Bruno and Sakmann 2006). Thus, activation of target cells at the subsequent processing stage can be assured by increasing either the rate or the synchronicity of discharges in the converging input connections. The advantage of increasing salience by synchronization is that integration intervals for synchronous inputs are very short, allowing for instantaneous detection of enhanced salience. Hence, information about the relatedness of responses can be read out very rapidly. In extremis, single discharges can be labelled as salient and identified as belonging to a particular assembly if synchronized with a precision in the millisecond range.

Again, however, it is not trivial to endow artificial recurrent networks with the dynamics necessary to solve the binding problem. It would require to implement oscillatory microcircuits and mechanisms ensuring selective synchronization of feature selective nodes. The latter, in turn, have to rely on Hebbian learning mechanisms for which there are yet no satisfactory hardware solutions. Hence, there are multiple reasons why the unique potential of recurrent networks is only marginally exploited by AI systems.

Computing in High-Dimensional State Space

Unlike contemporary AI systems that essentially rely on the deep learning algorithms discussed above, recurrent networks exhibit highly complex non-linear dynamics, especially if the nodes are configured as oscillators and if the coupling connections impose delays—as is the case for natural networks. These dynamics provide a very high-dimensional state space that can be exploited for the realization of functions that go far beyond those discussed above and are based on radically different computational strategies. In the following, some of these options will be discussed and substantiated with recently obtained experimental evidence.

The non-linear dynamics of recurrent networks are exploited for computation in certain AI systems, the respective strategies being addressed as “echo state, reservoir or liquid computing” (Lukoševičius and Jaeger 2009; Buonomano and Maass 2009; D’Huys et al. 2012; Soriano et al. 2013). In most cases, the properties of recurrent networks are simulated in digital computers, whereby only very few of the features of biological networks are captured. In the artificial systems the nodes act as simple integrators and the coupling connections lack most of the properties of their natural counterparts. They operate without delay, lack specific topologies and their gain is non-adaptive. Most artificial recurrent networks also lack inhibitory interneurons that constitute 20% of the neurons in natural systems and interact in highly selective ways with the excitatory neurons. Moreover, as the updating of network

states has to be performed sequentially according to the clock cycle of the digital computer used to simulate the recurrent network, many of the analogue computations taking place in natural networks can only be approximated with iterations if at all. Therefore, attempts are made to emulate the dynamics of recurrent networks with analogue technology. An original and hardware efficient approach is based on optoelectronics. Laser diodes serve as oscillating nodes and these are reciprocally coupled through glass fibres whose variable length introduces variations of coupling delays (Soriano et al. 2013). All these implementations have in common to use the characteristic dynamics of recurrent networks as medium for the execution of specific computations.

Because the dynamics of recurrent networks resemble to some extent the dynamics of liquids—hence the term “liquid computing”—the basic principle can be illustrated by considering the consequences of perturbing a liquid. If objects impact at different intervals and locations in a pond of water, they generate interference patterns of propagating waves whose parameters reflect the size, speed, location and the time of impact of the objects. The wave patterns fade with a time constant determined by the viscosity of the liquid, interfere with one another and create a complex dynamic state. This state can be analysed by measuring at several locations in the pond the amplitude, frequency and phase of the respective oscillations and from these variables a trained classifier can subsequently reconstruct the exact sequence and nature of the impacting “stimuli”. Similar effects occur in recurrent networks when subsets of nodes are perturbed by stimuli that have a particular spatial and temporal structure. The excitation of the stimulated nodes spreads across the network and creates a complex dynamic state, whose spatio-temporal structure is determined by the constellation of initially excited nodes and the functional architecture of the coupling connections. This stimulus-specific pattern continues to evolve beyond the duration of the stimulus due to reverberation and then eventually fades. If the activity has not induced changes in the gain of the recurrent connections the network returns to its initial state. This evolution of the network dynamics can be traced by assessing the activity changes of the nodes and is usually represented by time varying, high-dimensional vectors or trajectories. As these trajectories differ for different stimulus patterns, segments exhibiting maximal distance in the high-dimensional state space can be selected to train classifiers for the identification of the respective stimuli.

This computational strategy has several remarkable advantages: (1) low-dimensional stimulus events are projected into a high-dimensional state space where nonlinearly separable stimuli become linearly separable; (2) the high dimensionality of the state space can allow for the mapping of more complicated output functions (like the XOR) by simple classifiers, and (3) information about sequentially

presented stimuli persists for some time in the medium (fading memory). Thus, information about multiple stimuli can be integrated over time, allowing for the representation of sequences; (4) information about the statistics of natural environments (the internal model) can be stored in the weight distributions and architecture of the recurrent connections for instantaneous comparison with incoming sensory evidence. These properties make recurrent networks extremely effective for the classification of input patterns that have both spatial and temporal structure and share overlapping features in low-dimensional space. Moreover, because these networks self-organize and produce spatio-temporally structured activity patterns, they have generative properties and can be used for pattern completion, the formation of novel associations and the generation of patterns for the control of movements. Consequently, an increasing number of AI systems now complement the feed-forward strategy implemented in deep learning networks with algorithms inspired by recurrent networks. One of these powerful and now widely used algorithms is the Long Short Term Memory (LSTM) algorithm, introduced decades ago by Hochreiter and Schmidhuber (1997) and used in systems such as AlphaGo, the network that outperforms professional GO players (Silver et al. 2017, 2018). The surprising efficiency of these systems that excels in certain domains human performance has nurtured the notion that brains operate in the same way. If one considers, however, how fast brains can solve certain tasks despite of their comparatively extremely slow components and how energy efficient they are, one is led to suspect implementation of additional and rather different strategies.

And indeed, natural recurrent networks differ from their artificial counterparts in several important features which is the likely reason for their amazing performance. In sensory cortices the nodes are feature selective, i.e. they can be activated only by specific spatio-temporal stimulus configurations. The reason is that they receive convergent input from selected nodes of the respective lower processing level and thus function as conjunction-specific units in very much the same way as the nodes in feed forward multilayer networks. In low areas of the visual system, for example, the nodes are selective for elementary features such as the location and orientation of contour borders, while in higher areas of the processing hierarchy the nodes respond to increasingly complex constellations of elementary features. In addition, the nodes of natural systems, the neurons, possess an immensely larger spectrum of integrative and adaptive functions than the nodes currently used in artificial recurrent networks. And finally the neurons and/or their embedding microcircuits are endowed with the propensity to oscillate.

The recurrent connections also differ in important respects from those implemented in most artificial networks. Because of the slow velocity of signals conveyed by neuronal ax-

ons interactions occur with variable delays. These delays cover a broad range and depend on the distance between interconnected nodes and the conduction velocity of the respective axons. This gives rise to exceedingly complex dynamics and permits exploitation of phase space for coding. Furthermore and most importantly, the connections are endowed with plastic synapses whose gain changes according to the correlation rules discussed above. Nodes tuned to features that often co-occur in natural environments tend to be more strongly coupled than nodes responding to features that rarely occur simultaneously. Thus, through both experience-dependent pruning of connections during early development and experience-dependent synaptic plasticity, statistical contingencies between features of the environment get internalized and stored not only in the synaptic weights of feed-forward connections to feature selective nodes but also in the weight distributions of the recurrent connections. Thus, in low levels of the processing hierarchy the weight distributions of the recurrent coupling connections reflect statistical contingencies of simple and at higher levels of more complex constellations of features. In other words, the hierarchy of reciprocally coupled recurrent networks contains a model of the world that reflects the frequency of co-occurrence of typical relations among the features/components of composite perceptual objects. Recent simulation studies have actually shown that performance of an artificial recurrent network is substantially improved if the recurrent connections are made adaptive and can “learn” about the feature contingencies of the processed patterns (Lazar et al. 2009; Hartmann et al. 2015).

Information Processing in Natural Recurrent Networks

Theories of perception formulated more than a hundred years ago (von Helmholtz 1867) and a plethora of experimental evidence indicate that perception is the result of a constructivist process. Sparse and noisy input signals are disambiguated and interpreted on the basis of an internal model of the world. This model is used to reduce redundancy, to detect characteristic relations between features, to bind signals evoked by features constituting a perceptual object, to facilitate segregation of figures from background and to eventually enable identification and classification. The store containing such an elaborate model must have an immense capacity, given that the interpretation of ever-changing sensory input patterns requires knowledge about the vast number of distinct feature conjunctions characterizing perceptual objects. Moreover, this massive amount of prior knowledge needs to be arranged in a configuration that permits ultrafast readout to meet the constraints of processing speed. Primates perform

on average four saccades per second. This implies that new visual information is sampled approximately every 250 ms (Maldonado et al. 2008; Ito et al. 2011) and psychophysical evidence indicates that attentional processes sample visual information at comparable rates (Landau 2018; Landau and Fries 2012). Thus, the priors required for the interpretation of a particular sensory input need to be made available within fractions of a second.

How the high-dimensional non-linear dynamics of delay-coupled recurrent networks could be exploited to accomplish these complex functions is discussed in the following paragraph.

A hallmark of natural recurrent networks such as the cerebral cortex is that they are spontaneously active. The dynamics of this resting activity reflects the weight distributions of the structured network and hence harbours the entirety of the stored “knowledge” about the statistics of feature contingencies, i.e. the latent priors used for the interpretation of sensory evidence. This predicts that resting activity is high dimensional and represents a vast but constrained manifold inside the universe of all theoretically possible dynamical states. Once input signals become available they are likely to trigger a cascade of effects: They drive in a graded way a subset of feature sensitive nodes *and* thereby perturb the network dynamics. If the evidence provided by the input patterns matches well the priors stored in the network architecture, the network dynamics will collapse to a specific substate that provides the best match with the corresponding sensory evidence. Such a substate is expected to have a lower dimensionality and to exhibit less variance than the resting activity, to possess a specific correlation structure and be metastable due to reverberation among nodes supporting the respective substate. Because these processes occur within a very-high-dimensional state space, substates induced by different input patterns are usually well segregated and therefore easy to classify. As the transition from the high-dimensional resting activity to substates follows stimulus-specific trajectories, classification of stimulus-specific patterns is possible once trajectories have sufficiently diverged and long before they reach a fix point and this could account for the extremely fast operations of natural systems.

Experimental studies testing such a scenario are still rare and have become possible only with the advent of massive parallel recordings from the network nodes. So far, however, the few predictions that have been subject to experimental testing appeared to be confirmed. For the sake of brevity, these experimental results are not discussed here. They have been reviewed recently in Singer (2019a). A simplified representation of the essential features of a delayed coupled oscillator network supposed to be realized in the superficial layers of the cerebral cortex is shown in Fig. 3 (adapted from Singer 2018).

Fig. 3 Schematic representation of wiring principles in supra-granular layers of the visual cortex. The coloured discs (nodes) stand for cortical columns that are tuned to specific features (here stimulus orientation) and have a high propensity to engage in oscillatory activity due to the intrinsic circuit motif of recurrent inhibition. These functional columns are reciprocally coupled by a dense network of excitatory connections that originate mainly from pyramidal cells and terminate both on pyramidal cells and inhibitory interneurons in the respective target columns. Because of the genetically determined span of these

connections coupling decreases exponentially with the distance between columns. However, these connections undergo use-dependent selection during development and remain susceptible to Hebbian modifications of their gain in the adult. The effect is that the weight distributions of these connections and hence the coupling strength among functional columns (indicated by thickness of lines) reflect the statistical contingencies of the respective features in the visual environment (for further details see text). (From Singer W (2018) Neuronal oscillations: unavoidable and useful? *Europ J Neurosci* 48: 2389–2398)

Concluding Remarks

Despite considerable effort there is still no unifying theory of information processing in natural systems. As a consequence, numerous experimentally identified phenomena lack a cohesive theoretical framework. This is particularly true for the dynamic phenomena reviewed here because they cannot easily be accommodated in the prevailing concepts that emphasize serial feed-forward processing and the encoding of relations by conjunction-specific neurons. It is obvious, however, that natural systems exploit the computational power offered by the exceedingly complex, high-dimensional and non-linear dynamics that evolve in delay-coupled recurrent networks.

Here concepts have been reviewed that assign specific functions to oscillations, synchrony and the more complex dynamics emerging from a delay-coupled recurrent network and it is very likely that further computational principles are realized in natural systems that wait to be uncovered. In view of the already identified and quite remarkable differences between the computational principles implemented in artificial and natural systems it appears utterly premature to enter discussions as to whether artificial systems can acquire functions that we consider proper to natural systems such as intentionality and consciousness or whether artificial

agents can or should be considered as moral agents that are responsible for their actions. Even if we had a comprehensive understanding of the neuronal underpinnings of the cognitive and executive functions of human brains—which is by no means the case—we still would have to consider the likely possibility, that many of the specific human qualities cannot be deduced from the neuronal functions of individual brains alone but owe their existence to cultural evolution. As argued elsewhere (Singer 2019b), it is likely that most of the connotations that we associate with intentionality, responsibility, morality and consciousness are attributions to our self-model that result from social interactions of agents endowed with the cognitive abilities of human beings. In a nutshell the argument goes as follows: Perceptions—and this includes also the perception of oneself and other human beings—are the result of constructivist processes that depend on a match between sensory evidence and a-priori knowledge, so-called priors. Social interactions between agents endowed with the cognitive abilities of humans generate immaterial realities, addressed as social or cultural realities. This novel class of realities assume the role of implicit priors for the perception of the world and oneself. As a natural consequence perceptions shaped by these cultural priors impose a dualist classification of observables into material and immaterial phenomena, nurture the concept of ontological substance dualism and generate the epistemic conundrum of experiencing oneself

The Emergence of Qualia

Fig. 4 Phase transitions during biological (left) and cultural evolution (right) that lead to the emergence of new qualities. For details see text. (From Singer W (2019) A naturalistic approach to the hard problem of consciousness. *Front Syst Neurosci* 13: 58)

as existing in both a material and immaterial dimension. Intentionality, morality, responsibility and certain aspects of consciousness such as the qualia of subjective experience belong to this immaterial dimension of social realities.

This scenario is in agreement with the well-established phenomenon that phase transitions in complex systems can generate novel qualities that transcend the qualities of the systems' components. The proposal is that the specific human qualities (intentionality, consciousness, etc.) can only be accounted for by assuming at least two-phase transitions: One having occurred during biological evolution and the second during cultural evolution (Fig. 4). The first consists of the emergence of cognitive and executive functions from neuronal interactions during biological evolution and the second of the emergence of social realities from interactions between the cognitive agents that have been brought forth by the first phase transition. Accordingly, different terminologies (*Sprachspiel*) have to be used to capture the qualities of the respective substrates and the emergent phenomena, the neuronal interactions, the emerging cognitive and executive functions (behaviour), the social interactions among cognitive agents and the emerging social realities. If this evolutionary plausible scenario is valid, it predicts, that artificial agents, even if they should one day acquire functions resembling those of individual human brains,—and this is not going to happen tomorrow—will still lack the immaterial dimensions of our self-model. The only way to acquire this dimension—at least as far as I can see—would be for them

to be raised like children in human communities in order to internalize in their self-model our cultural achievements and attributions—and this would entail not only transmission of explicit knowledge but also emotional bonding. Or these man-made artefacts would have to develop the capacity and be given the opportunity to engage in their own social interactions and recapitulate their own cultural evolution.

References

- Barlow, H. B. (1972). Single units and sensation: A neurone doctrine for perceptual psychology? *Perception*, 1, 371–394.
- Bellec, G., Scherr, F., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2019). Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. *arXiv 2553450*, 2019, 1–34.
- Bruno, R. M., & Sakmann, B. (2006). Cortex is driven by weak but synchronously active thalamocortical synapses. *Science*, 312, 1622–1627.
- Buonomano, D. V., & Maass, W. (2009). State-dependent computations: Spatiotemporal processing in cortical networks. *Nature Reviews. Neuroscience*, 10, 113–125.
- D'Huys, O., Fischer, I., Danckaert, J., & Vicente, R. (2012). Spectral and correlation properties of rings of delay-coupled elements: Comparing linear and nonlinear systems. *Physical Review E*, 85(056209), 1–5.
- DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11, 333–341.
- Felleman, D. J., & van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1, 1–47.
- Frey, U., & Morris, R. G. M. (1997). Synaptic tagging and long-term potentiation. *Nature*, 385, 533–536.
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkin-

- son, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536, 171–178.
- Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338, 334–337.
- Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35, 96–111.
- Hahn, G., Petermann, T., Havenith, M. N., Yu, Y., Singer, W., Plenz, D., & Nikolic, D. (2010). Neuronal avalanches in spontaneous activity in vivo. *Journal of Neurophysiology*, 104, 3312–3322.
- Hartmann, C., Lazar, A., Nessler, B., & Triesch, J. (2015). Where's the noise? Key features of spontaneous activity and neural variability arise through learning in a deterministic network. *PLoS Computational Biology*, 11(12), e1004640, 1–35.
- Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley & Sons.
- Hirabayashi, T., Takeuchi, D., Tamura, K., & Miyashita, Y. (2013). Microcircuits for hierarchical elaboration of object coding across primate temporal areas. *Science*, 341, 191–195.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780.
- Hopfield, J. J. (1987). Learning algorithms and probability distributions in feed-forward and feed-back networks. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 8429–8433.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195, 215–243.
- Ito, J., Maldonado, P., Singer, W., & Grün, S. (2011). Saccade-related modulations of neuronal excitability support synchrony of visually elicited spikes. *Cerebral Cortex*, 21, 2482–2497.
- Landau, A. N. (2018). Neuroscience: A mechanism for rhythmic sampling in vision. *Current Biology*, 28, R830–R832.
- Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, 22, 1000–1004.
- Lazar, A., Pipa, G., & Triesch, J. (2009). SORN: A self-organizing recurrent neural network. *Frontiers in Computational Neuroscience*, 3(23), 1–9.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444.
- Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3, 127–149.
- Maldonado, P., Babul, C., Singer, W., Rodriguez, E., Berger, D., & Grün, S. (2008). Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *Journal of Neurophysiology*, 100, 1523–1532.
- Milner, P. M. (1992). The functional nature of neuronal oscillations. *Trends in Neurosciences*, 15, 387.
- Plenz, D., & Thiagarajan, T. C. (2007). The organizing principles of neuronal avalanches: Cell assemblies in the cortex? *Trends in Neurosciences*, 30, 99–110.
- Quiñ Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435, 1102–1107.
- Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: The synaptic tagging and capture hypothesis. *Nature Reviews. Neuroscience*, 12, 17–30.
- Rosenblatt, F. (1958). The perceptron. A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386–408.
- Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362, 1140–1144.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, 550, 354–359.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, 24, 49–65.
- Singer, W. (2018). Neuronal oscillations: Unavoidable and useful? *The European Journal of Neuroscience*, 48, 2389–2398.
- Singer, W. (2019a). Cortical dynamics. In J. R. Lupp (Series Ed.) & W. Singer, T. J. Sejnowski, & P. Rakic (Vol. Eds.), Strüngmann Forum reports: Vol. 27. The neocortex. Cambridge, MA: MIT Press (in print). 167–194.
- Singer, W. (2019b). A naturalistic approach to the hard problem of consciousness. *Frontiers in Systems Neuroscience*, 13, 58.
- Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 18, 555–586.
- Soriano, M. C., Garcia-Ojalvo, J., Mirasso, C. R., & Fischer, I. (2013). Complex photonics: Dynamics and applications of delay-coupled semiconductor lasers. *Reviews of Modern Physics*, 85, 421–470.
- Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311, 670–674.
- Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews. Neuroscience*, 5, 97–107.
- von der Malsburg, C., & Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, 67, 233–242.
- von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Hamburg, Leipzig: Leopold Voss Verlag.
- Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., O'Shea, D. J., Sohal, V. S., Goshen, I., Finkelstein, J., Paz, J. T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J. R., Hegemann, P., & Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, 477, 171–178.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.