Computational neuroscience has focused largely on the dynamics and function of local circuits of neuronal populations dedicated to a common task, such as processing a common sensory input, storing its features in working memory, choosing between a set of options dictated by controlled experimental settings or generating the appropriate actions. Most of current circuit models suggest mechanisms for computations that can be captured by networks of simplified neurons connected via simple synaptic weights. In this article I review the progress of this approach and its limitations. It is argued that new experimental techniques will yield data that might challenge the present paradigms in that they will (1) demonstrate the computational importance of microscopic structural and physiological complexity and specificity; (2) highlight the importance of models of large brain structures engaged in a variety of tasks; and (3) reveal the necessity of coupling the neuronal networks to chemical and environmental variables.

Addresses
1 Edmond and Lily Safra Center for Brain Sciences, The Hebrew University, Jerusalem 91904, Israel
2 Center for Brain Science, Harvard University, Cambridge, MA 02138, USA

Corresponding author: Sompolinsky, Haim (haim@fiz.huji.ac.il)

Introduction
The last decade witnessed impressive expansion in the scope of the systems and questions studied by theoretical neuroscience. Models became increasingly more complex and more realistic, and analysis methods more sophisticated. These theoretical investigations advanced the understanding of the function of neuronal circuits in several ways. Studies of the neural code developed quantitative measures of the processing accomplished by early sensory stages. Comparing the neural code with bounds set by optimality criteria revealed underlying design principles. Other studies focused on structural and dynamic mechanisms for computations performed by specific local circuits, including sensory processing, working memory, decision-making, motor control, and neural learning and memory. In most of these models, computation is the outcome of appropriate patterns of connectivity, rather than complex cellular or molecular properties; the latter are considered as determining the few 'control parameters' that govern the functionality of the entire circuit.

In this article, I will review some of the progress made in past decades and will highlight some of the limitations in present approaches. I will suggest directions for extending the scope of current theories of neuronal computations in the context of recent and future experimental efforts in systems neuroscience.

Neural codes and neural mechanisms
The study of the neural code is perhaps the most developed field of theoretical neuroscience. Information theory and advanced statistical inference methods provide new frameworks for quantitative assessments of the sensory information represented and processed by neuronal systems from single neurons and synapses to large populations [1–3]. By comparing these results with psycho-physical performance, one gains insight into the selective contributions of specific circuits to the behavioral outcome. At a higher level, efficient coding theory addresses the why question. Using a variety of optimality criteria, the theory explores the conditions under which observed performance at both the neuronal and behavioral levels is close to optimal. Optimality theories have successfully explained salient features of the neural code in primary sensory systems. In particular they have elucidated the forms of the receptive fields and their adaptive properties, as well as the underlying constraints such as the sources of noise, and wiring and metabolic costs [4,5]. These theories highlight the statistical nature of information processing and the importance of testing performance using stimulus ensembles that closely mimic the natural environment of the organism. Recently, Bayesian optimality theories have been developed for cognition and behavior [6] (see also [7]), but at present, their neuronal correlates are unclear [8].

The central goal of theoretical neuroscience remains to answer the how question, namely to elucidate the mechanisms underlying brain function, linking structure, dynamics, and computation in neuronal circuits. Translating hypotheses into well-defined mathematical models that incorporate the relevant details of the studied system allows their scrutiny by mathematical analysis and simulations, and provides a framework for making predictions for further experimental testing. While theoretical neuroscience has generated proposals of interesting and sometimes profound potential mechanisms for many brain
functions, proving that these mechanisms are actually implemented by the brain, or rejecting this scenario has been difficult. This is partly because of the lack of sufficient information about the underlying circuitry, and partly because the brain often uses a mixture of mechanisms and pathways. Hopefully, new technologies such as multi-electrode recordings, optogenetics and connectomics will provide highly specific probes and perturbation methods that will enhance significantly the ability to test and revise proposed mechanisms.

Detailed theoretical investigations may reveal the limitations of potential mechanisms, reducing the likelihood of their relevance to the real brain. One example is the Hopfield model for associative memory [9]. While for random memorized patterns the theoretical estimate of the model’s memory capacity scales reasonably with the number of synapses, the theory revealed a dramatic drop in capacity when memorized patterns are structured, strongly suggesting that the circuits for long term memory differ from the model either in their architecture or in their learning rule [10]. A Hopfield network exhibits a discrete set of fixed-point attractors. Other circuits exhibit continuous manifolds of stable states called attractor manifolds. Such manifolds have been proposed in the context of sensory processing [11], spatial working memory [12,13], neural integrators [14], and spatial navigation systems [15]. While the theory demonstrates the functional utility of attractor manifolds, such as line and ring attractors, it also predicts that they require fine-tuning of certain critical parameters [16,17]. Interestingly, in some cases (e.g., grid cells [18], oculomotor integrator, the head direction system), these mechanisms seem the only feasible ones, suggesting that the presence of a robust (as yet unknown) mechanism of maintaining the relevant critical parameters at their required values.

The network paradigm: theory of everything?
The recent massive increase in the power of low cost computers allowed the simulations of larger and more realistic neuronal networks. In addition, theoretical analysis yielded increasingly sophisticated and finessed understanding of the dynamics and information processing of circuits with complex features such as spiking neurons, short-term synaptic plasticity, and multiple types of synaptic conductance. Nevertheless, with a few notable exceptions [19–21], the majority of present theories continue to adhere to the network paradigm, which attributes the computing power of neuronal systems mostly to the clever patterns of connectivity, rather than to the complexity of the individual ‘nodes’ and connections. Thus, receptive fields are determined by specific patterns of feedforward connections; the balance between excitatory and inhibitory feedback, as well as the specific topographic and functional patterns of these connections, shape computation performed by recurrent networks; and information acquired through learning and memory is stored in the strength of individual synapses (the synaptic efficacies). In many cases, it has been demonstrated that the function of realistic circuits remains largely intact when they are reduced to rate-based dynamics with scalar input-output nonlinearity, connected by simple synaptic ‘weights’. Reduced network models of N such units are typically similar in form to the following equations,

$$\tau_i \frac{d r_i}{dt} = -r_i + g \left( \sum_{j=1}^{N} w_{ij} r_j + I_i \right)$$

(1)

where $r_i$ and $I_i$ are, respectively, the time constant and the external input of the $i$th neuron, $w_{ij}$ is the synaptic connection matrix between pairs of neurons, and $g$ is the nonlinear function relating the neuron’s synaptic input at time $t$, and its output rate variable, $r_i(t)$. Slightly more complex models describe networks of interconnected Integrate-and-Fire ‘point neurons’, which incorporate discrete spiking events at threshold crossings of the membrane potential, and possibly including synaptic conductances and simplified short-term synaptic plasticity [22]. Clearly, such models do not capture many of the salient features of brain dynamics, most notably, its rich cohort of rhythmic patterns, which characterize the different global states of the brain [23,24]. However, their role in ongoing information processing remains unknown.

Will the solution of a set of differential equations governing the dynamics of a network of 100 billion simple units with architecture similar to the human brain and driven by the appropriate sensory input, exhibit something resembling thought, visual cognition, language, or consciousness? There are several reasons to be cautious. First, the demonstrated computing capabilities of current network models are still far from that of the brain. To cite a few examples, we do not have yet plausible network models for language processing, for representing and manipulating conceptual knowledge, for flexible learning, for high level perceptual or social cognition, or for planning and monitoring complex behavior. Thus, we cannot rule out the possibility that to achieve more powerful performance, models will have to incorporate the vast richness of the structural and physiological properties of real neurons and synapses.

A fundamental deficiency of the network paradigm is the representation of the environment in terms of a set of external inputs. In natural behaving scenarios, the brain is not a passive receiver of sensory sensations, but actively seeks information, implying that a full theory of brain function must incorporate the dynamic interaction between the neuronal network and the environment, from the organism’s sensors and actuators, through its navigation in the environment, to its interactions with other animals [20,25,26].
A similar problem exists with regards to the brain’s molecular environment. Numerous molecular processes control brain electrical signaling [21]. For instance, neuro-modulatory transmitters control the global brain state but also affect specific circuits and functions, and in turn are partially controlled by neuronal activity [19]. These signals propagate often via volume transmission, rather than one-to-one synaptic contacts, some are mediated by glial cells; hence their dynamics is not captured by the network synaptic communications. Some of these modulatory effects are slow relative to the time scale of the electrical signaling, allowing one to lump these effects into the static structure of the network, in a separation-of-time-scale approach. However, the range of time scales of some of these processes overlaps with the spectrum of time scales of slow synaptic receptors such as NMDA and short-term synaptic plasticity. The reciprocal dynamic interaction between electrical and chemical degrees of freedom implies that the latter cannot be fully captured by simply representing them as external inputs or other network parameters.

The abundance of action-perception cycles and electrochemical loops imply that the neuronal networks should be viewed as part of a large ensemble of interacting networks of vast complexity.

Lastly, incorporating learning, a key factor in the computational power of the nervous system, requires modeling the plasticity dynamics of synapses. A minimal approach would add to Eq. (1) a set of rules for experience dependent changes in the $w_{ij}$ in the general form of

$$\frac{dw_{ij}}{dr} = f(r_i, r_j, w_{ij}, e_{ij})$$

(2)

where $e_{ij}$ is an input that carries information about reward and other types of feedback from the environment or from other brain systems. More refined versions (STDP rules) include dependences on the spike times of the presynaptic and postsynaptic neurons [27]. At the moment, we do not have a concrete version of these equations that exhibits the power of learning of the nervous system. Translating some of the impressive advances in machine learning algorithms to neuronal networks has been successful, so far, only in simple architectures and tasks [28]. Most importantly, $e_{ij}$ reflects the interaction between the synaptic dynamics and the external as well as the biochemical environment. Molecular and cellular studies have begun to map the vast molecular pathways underlying synaptic plasticity [21,29]. From the present perspective, the important issue is whether, at the circuit level, these molecular computations can be abstracted in terms of specific forms of plasticity rules, such as Eq. (2), or need to be accounted for as additional dynamical degrees of freedom, which would again imply that a circuit-level theory based only on the electrical degrees of freedom is inadequate.

**Single circuits and single functions**

Given the enormous challenges associated with an attempt to understand brain function in its totality, it is no wonder why most of neuroscience research, experimental as well as theoretical, focuses on understanding subfunctions measured in neuronal subsystems in carefully controlled environments. This parcellation of brain function and structure is done largely on practical bases, raising questions about which are the natural computational modules of the nervous system. Anatomically discrete structures such as neurons, synapses, and nuclei, are natural candidates. The local functional organization and connectivity of many systems suggest additional candidates, such as a patch of retina of a (linear) size of a few hundred microns, a ‘barrel’ column in rodents S1, spanning a few hundred microns in cortical surface, or a 1 mm patch of V1 (called a cortical hypercolumn). While some of these circuits are part of a continuum, they are singled out as functional modules in that their neurons are co-activated by the same localized sensory stimulus.

Theoretical research over the past decades made significant headway in understanding the dynamic repertoire of local neuronal circuits, primarily in cortex, and its functional consequences. Circuits dominated by excitatory synaptic recurrence may exhibit sustained signal amplification, multiple attractors and attractor manifolds [30]. Dominant recurrent inhibition may give rise to enhanced selectivity, sparse coding [31], signal normalization [32], inhibitory-mediated transient amplification [33], and balanced states [34]. In addition, randomly connected circuits may give rise to complex, slow dynamics, including chaos [35], and have been recently proposed as cortical generators of spatio-temporal signals [36]. Applications of these properties include sensory feature selectivity, working and long-term memory, spatial navigation, decision-making, and motor control. Learning models have been applied to explain rule learning, adaptive behavior, and motor learning such as birdsong [28,37,38]. Common to most studies is the focus on a (loosely defined) local circuit, consisting of a population of $10^3-10^6$ neurons responding to the same stimulus and participating in solving a common task.

Experiments on neuronal functions focus on single neurons and local neuronal populations probed with highly constrained sets of stimuli and tasks. These studies often lead to the assignment of simple roles to single neurons in terms of their receptive or motor fields, or preferred stimuli and actions. Experiments using more complex stimuli and tasks reveal a plethora of nonlinear, contextual, and even cross-modal effects in particular in behaving animals [39–43]. These provide strong evidence that the notion of local processing in cortex and other structures has restricted utility, which is consistent with the abundance of long-range cortico-cortical connections and interactions through subcortical structures. Thus, it is
not surprising that concepts and dynamic models based on the notion of purely local processing are difficult to match against experimental data.

**Beyond the local circuits**

New methods for imaging and recording from multiple areas simultaneously as well as new virtual reality methods for behavioral paradigms open new possibilities for observing brain activity at large spatial and temporal scales and in natural environments. Data accumulated with these methods pose new challenges for theoretical neuroscience, including the task of replacing concepts and models based on local circuits dedicated to single functions, by frameworks suitable for dealing with computations in integrated large-scale systems performing multiple computations.

Efforts to meet this challenge will benefit from several encouraging developments. On the theoretical side, spatial averaging methods [44] based on statistical physics, and temporal averaging [45] based on dynamic theory, have been successful in bridging the scale from single neurons to homogenous populations. Development of more sophisticated multi-scaling techniques, similar to those developed in material science and fluid dynamics [46,47], may help in developing the theory of the dynamics of larger, more heterogeneous neuronal circuits. Multi-scaling methods will also allow more systematic determination of which microscopical details are sufficient for the understanding of a given set of computations.

While efficient coding theory has been largely confined to early sensory stages, recent models of object recognition in vision have begun to shed light on how cascades of sensory processing stages can modify the ‘format’ of sensory representation, making it readily accessible by downstream systems [48]. The recent advances in machine learning applications of Deep Networks [49,50] provide new evidence for the computing power of multilayer systems, although devising biologically plausible learning algorithms for these systems remains an important challenge. Other interesting large-scale studies include generative models, predictive coding, and Bayesian networks [51–53]. Neuroscientists may also gain useful lessons from the design principles adopted in the construction of autonomous agents and intelligent robotic systems [54,55].

Several works suggest ways in which neuronal systems can multiplex a range of tasks and efficiently and rapidly store and read embedded information. Valiant [56] has shown how appropriate graph properties of sparse cortical networks may enable the solution of multiple random access tasks including forming representations of new concepts, establishing relations between existing concepts, and learning threshold linear rules. These tasks may be considered primitives of cognitive abilities [56]. Neuronal global workspace models studied by Dehaene and colleagues [57,58] propose architectures that enable conscious processing, i.e., the routing and sharing of information, as well as nonlinear amplification and selections of signals, across large networks of interacting local modules. Eliasmith and Trujillo [59] have explored the implementation of integrated cognitive functions in large-scale brain simulations.

Computational neuroscience of disease is a new growth area of multi-scale research. The wide dissatisfaction from DSM-type classifications of mental disorders [60], the realization of the multiplicity and complexity of the genetic and molecular causal chains leading to specific diseases, and the discovery that many of these pathways are shared by clusters of diseases, drew attention to systems approaches in disease research, triggering new statistical and modeling efforts [61,62]. Indeed several models have begun to address the circuit dysfunctions associated with neurological and psychiatric diseases, for example epilepsy, schizophrenia, autism, Parkinson’s disease, and disorders of consciousness [23,24,63]. These models typically link cellular and molecular events to aberrations in the local circuit dynamics (e.g., abnormal oscillations, altered gains of neurons and synapses, and excitation-inhibition imbalance) and then to changes in patterns of large-scale brain dynamics, ultimately leading to cognitive deficits.

Theories of Reinforcement Learning have been successfully applied to data from human imaging and multi-site recording in primates and rodents, shedding light on mechanisms that link sensory processing, decision-making and reward, in normal and in pathological conditions [64], and making an important contribution to computational psychiatry [65].

Structural and functional imaging of the human brain, including the discovery of resting state networks [66], opened new vistas to the large-scale organization of the human brain and its relations to cognitive functions and states. These and data from nonhuman primates led to important computational work on graph theoretic analysis of large-scale neuronal systems, as well as models that probe spatio-temporal activity patterns in these systems [67–69].

Finally, advances in experimental techniques have led to research initiatives designed to densely map the structure and function of entire sensory-motor pathways as well of the whole nervous systems of simple organisms, such as the nematode worm *C. elegans* [70–72], and larval zebra fish [73]. These endeavors will provide excellent model systems for developing and testing theories of the dynamics and computations of integrated neuronal systems. Recordings of the behaviors of these animals in
realistic conditions across long time spans will hopefully also elucidate the nature of the interaction between the neuronal networks and the chemical and sensory-motor environments.

Conclusions
The new technologies that appear in the horizon of brain research, promise to yield information about the structure and function of neuronal circuits with unprecedented detail and to allow their manipulations with exquisite specificity in real time within realistic behavioral settings. These exciting developments together with advances in human brain structural and functional imaging, present new opportunities and challenges for theoretical neuroscience. New databases can be used to study biologically realistic circuit models and test the relevance of the microscopic complexity to computations at the level of local circuits. These studies might demonstrate that the molecular specificity of cell and synaptic types with their complex biophysics plays a major role in neuronal computations.

Data from recordings and imaging of entire pathways in complex sensory and behavioral environments will require expansion of our theories and models to include integrated heterogeneous neuronal systems that can multiplex a spectrum of sensory, motor and cognitive tasks. Mapping of the modulatory networks of the brain and the perception-action loops will challenge the currently dominant computational paradigm that focuses entirely on electrical degrees of freedom.

It will be interesting to see if future developments will lead not only to more accurate neuronal network models but also to a paradigm shift in which the neuronal circuits are considered part of a web of interacting molecular and environmental networks that together shape brain function. If this turns out to be the case, then brain theory must develop new sorts of abstractions for the ‘embedded’ system. Ultimately, the success of brain theory hinges upon its ability to carve nature at its joints.

Acknowledgements
HS thanks Kenneth Blum for helpful comments on the manuscript. Work is supported in part by Gatsby Charity Foundation and Max Planck – Hebrew University Center for Neurobiology.

References
6 Perspective


