

Theoretical Neuroscience Rising

L.F. Abbott^{1,*}

¹Department of Neuroscience and Department of Physiology and Cellular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

*Correspondence: lfabbott@columbia.edu

DOI 10.1016/j.neuron.2008.10.019

Theoretical neuroscience has experienced explosive growth over the past 20 years. In addition to bringing new researchers into the field with backgrounds in physics, mathematics, computer science, and engineering, theoretical approaches have helped to introduce new ideas and shape directions of neuroscience research. This review presents some of the developments that have occurred and the lessons they have taught us.

Introduction

Twenty years ago, when *Neuron* got its start, theoretical neuroscience was experiencing a start of its own. Of course, there were important theoretical contributions to neuroscience long before 1988, most notably: the development of what we now call the integrate-and-fire model by Lapicque in 1907; the modeling of the action potential by Hodgkin and Huxley, a brilliant theoretical offshoot of their experimental work; the development of dendritic and axonal cable theory by Wilfred Rall; and the broad insights of David Marr. Nevertheless, over the past 20 years, theoretical neuroscience has changed from a field practiced by a few multitalented experimentalists and dedicated theorists (Jack Cowan, Steven Grossberg, John Rinzel, and Terry Sejnowski being early examples) sparsely scattered around the world to an integral component of virtually every scientific meeting and major department. Something has changed. How did this happen, and what impact has it had?

Two developments in the mid-1980s set the stage for the rapid expansion of theoretical neuroscience. One was the popularization of the backpropagation algorithm for training artificial neural networks (Rumelhart and McClelland, 1986). This greatly expanded the range of tasks that artificial neural networks could perform and led to a number of people entering neural network research. Around the same time, Amit, Gutfreund, and Sompolinsky (Amit et al., 1985) showed how a memory model proposed by Hopfield (1982) could be analyzed using methods of statistical physics originally designed for spin glasses. The sheer beauty of this calculation drew a large batch of physicists into the field. These new immigrants entered with high confidence-to-knowledge ratios that, hopefully, have been reduced through large growth in the denominators and more modest adjustments of the numerators.

What has a theoretical component brought to the field of neuroscience? Neuroscience has always had models (how would it be possible to contemplate experimental results in such complex systems without a model in one's head?), but prior to the invasion of the theorists, these were often word models. There are several advantages of expressing a model in equations rather than words. Equations force a model to be precise, complete, and self-consistent, and they allow its full implications to be worked out. It is not difficult to find word models in the conclusions sections of older neuroscience papers that sound reasonable but, when expressed as mathematical models, turn out to

be inconsistent and unworkable. Mathematical formulation of a model forces it to be self-consistent and, although self-consistency is not necessarily truth, self-inconsistency is certainly falsehood.

A skillful theoretician can formulate, explore, and often reject models at a pace that no experimental program can match. This is a major role of theory—to generate and vet ideas prior to full experimental testing. Having active theoretical contributors in the field allows us collectively to contemplate a vastly greater number of solutions to the many problems we face in neuroscience. Both theorists and experimentalists generate and test ideas, but due to the more rapid turnover time in mathematical and computational compared to experimental analyses, theorists can act as initial filters of ideas prior to experimental investigation. In this regard, it is the theorist's job to develop, test, frequently reject, and sometimes promote new ideas.

Theoretical neuroscience is sometimes criticized for not making enough predictions. This is part of a pre-versus-post debate about the field that has nothing to do with synapses. Although there are notable examples of predictions made by theorists and later verified by experimentalists in neuroscience, examples of postdictions are far more numerous and often more interesting. To apply prediction as the ultimate test of a theory is a distortion of history. Many of the most celebrated moments in quantitative science—the gravitational basis of the shape of planetary orbits, the quantum basis of the spectrum of the hydrogen atom, and the relativistic origin of the precession of the orbit of Mercury—involved postdictions of known and well-characterized phenomena. In neuroscience especially, experimentalists have gotten a big head start. There is nothing wrong with a model that “postdicts” previously known phenomena. The key test of the value of a theory is not necessarily whether it predicts something new, but whether it makes postdictions that generalize to other systems and provide valuable new ways of thinking.

The development of a theoretical component to neuroscience research has had significant educational impact across the biological sciences. The Sloan-Swartz initiative, for example, has supported almost 80 researchers who successfully transitioned from other fields to faculty positions in neuroscience. Jim Bower and Christof Koch set up the computational neuroscience course at Woods Hole, a summer course that is still educating people with backgrounds in both the biological and physical sciences and that has been copied in courses around the world.

Biology used to be a refuge for students fleeing mathematics, but now many life sciences students have a solid knowledge of basic mathematics and computer programming, and those that don't at least feel guilty about it. A number of developments have led to this shift, the rise of theoretical neuroscience certainly being one of them.

The following sections provide a sparse sampling of theoretical developments that have occurred over the past 20 years and discuss some of the things they have taught us. The presentation is idiosyncratic, with some developments presented in a different context than when they first appeared and perhaps from what their creators intended, and many important achievements ignored entirely. The focus is on lessons learned from a subset of the theoretical advances over the past 20 years.

Basic Principles

Many researchers have sought basic principles to help guide us through the complexities of neural circuits and cognition. Examples, discussed in the following paragraphs, are efficient coding, Bayesian inference, generative models, causality, and what I call the positivity of the neural code. The last of these, perhaps more accurately termed a basic constraint, is considered in a bit more detail to highlight a number of different developments.

The efficient coding hypothesis, formulated by Horace Barlow (Barlow, 2001), postulates that sensory systems are adapted to convey information about natural stimuli faithfully using a minimal amount of activity. This idea has been used to account for the receptive field properties of neurons in the retina (Atick and Redlich, 1990) and lateral geniculate nucleus (Dong and Atick, 1995), in primary visual cortex (Olshausen and Field, 1996; Bell and Sejnowski, 1997), in the fly visual system (van Hateren, 1997; Niven et al., 2007), and in the auditory system (Lewicki, 2002). An offshoot of this work has been a better understanding of the statistical properties of natural stimuli, such as visual scenes (Field, 1987; Tolhurst et al., 1992; Ruderman and Bialek, 1994; van der Schaaf and van Hateren, 1996).

The application of Bayesian inference to both neural systems and the behaviors they generate is another "first principles" approach to neuroscience. As an example of Bayesian inference, suppose that our faith in a certain statement S being true is characterized by a probability $P(S)$. Now, imagine that we make an observation O that occurs with probability $P(O|S)$ if S is true and with probability $P(O)$ whether or not S is true. Bayesian inference says that our belief in the veracity of the statement S after this observation should be $P(S)P(O|S)/P(O)$. The key point here is that our belief in S upon observation of O should change by an amount that depends on the how much more likely the truth of S makes O , quantified by the ratio $P(O|S)/P(O)$. Suppose that S is the statement that a pair of dice is loaded, and O is the observation of several consecutive rolls of double 6. This observation clearly increases the probability that S is true and, according to Bayesian inference, it should do so to the extent that the loaded dice hypothesis makes those rolls more likely. Bayesian inference thus provides a principle for quantifying the effect that evidence should have on belief (specifically how new evidence and prior expectation should be combined), and it can be used to derive optimal ways that multiple forms of evidence should be weighed in making a decision. On the behavioral level, Bayes-

ian inference has been applied to quantify how subjects combine multiple sources of information in proportion to their reliability (Ernst and Banks, 2002) and to account for aspects of motor control (Kording and Wolpert, 2006) and perception (Knill and Richards, 1996; Stocker and Simoncelli, 2006). In addition, proposals have been made for how such computations might be performed at the neuronal level (Ma et al., 2006).

A major problem we face in contemplating memory storage or perceptual processing is that we do not understand how neural activity represents information beyond fairly early stages in sensory processing. Generative models provide a guiding principle for thinking about higher-order neural representations (Hinton and Ghahramani, 1997; Rao et al., 2002). In a generative model, patterns of neural activity in a high-level area must not only represent the data, they must also be capable of generating patterns of activity at earlier sensory stages, through back-projections, that resemble the activity evoked by the external world. Representations with such generative capabilities provide a good basis for constructing networks that perform complex tasks (see, for example, Hinton et al., 2006).

Causality, the temporal relationship between cause and effect, is a basic fact of life that has important scientific implications, quantified in physics by the Kramers-Kronig relations. In neuroscience, we have Hebb's rule of synaptic plasticity and its likely biophysical substrate, spike-timing-dependent plasticity, or STDP (Dan and Poo, 2006). Hebb stated that a synapse should be strengthened when it participates in *causing* a postsynaptic neuron to fire, and STDP not only realizes this rule and its causality condition, it also leads to weakening of synapses when there is correlated pre- and postsynaptic activity that is not causally ordered in time. The theoretical implications of STDP have occupied and continue to engage theorists across a wide range of applications (for example, Gerstner et al., 1996; Song and Abbott, 2001; Buchs and Senn, 2002; Nowotny et al., 2003; Morrison et al., 2007; Izhikevich, 2007; Kang et al., 2008).

Spike counts and neuronal firing rates are positive quantities. This simple fact has important implications for neural coding and neural circuits that provide a framework for thinking about a number of research directions taken over the past 20 years. Suppose that a set of different events results in various numbers of spikes being delivered to a neuron over a short period of time. The incoming action potentials appear on N separate presynaptic afferents. With such an arrangement, we would expect the number of incoming action potentials averaged across events to be proportional to N , because each input contributes a positive number to the total. Being an average over all events, this order- N input contains no information about the identity of any individual event. Such identity is contained in the event-to-event variance of each input about its average value. Note that this event-related variance is not noise; it is the signal that provides information about the distinctive properties of each event. If we assume that the inputs react independently to the different events, we expect these variances to add, producing a total event-related variance that is, like the average, proportional to N . Because the variance is a squared quantity, this means that the amplitude of the fluctuations that provide information about event identity is only proportional to \sqrt{N} . Thus, the ratio of the component of the total input that is useful for discriminating

between different events to the average size of the input is $\sqrt{N}/N = 1/\sqrt{N}$. For N of order 10,000, this ratio is only 1%. In other words, the large number of inputs to a neuron, which is a key feature of most neural circuits, appears in this light to be a liability rather than an advantage. When integrating a large number of synapses, the information-carrying fluctuations in input spiking can get drowned out by the noninformative din provided by the average rate of input spiking. A number of features of neuronal circuitry that have been studied by both experimental and theoretical neuroscientists over the past 20 years can be seen as responses to this basic problem.

How can this problem be solved? Proposed solutions can be divided into three broad classes: (1) correlate the inputs, (2) reduce N , and (3) make use of a balance between excitation and inhibition to cancel the noninformative component of the total input. Correlating the inputs means modifying the assumption that the different inputs to a neuron are independent. If we take the extreme case when, instead, all the inputs carry the same information, the problem is solved because the event-related variance is of order N^2 rather than N . This makes the ratio of informative to noninformative signal of order 1 rather than $1/\sqrt{N}$. Solutions of this form are seen at early stages of sensory processing, most dramatically in olfaction, where receptors with similar response characteristics project to common targets. Away from the periphery, the idea can still be applied by using correlation or synchrony to boost the efficacy of informative sets of inputs. Correlated or synchronous presynaptic action potentials have a greater impact on the response of a postsynaptic neuron than asynchronous spikes (Salinas and Sejnowski, 2001; Womelsdorf and Fries, 2007). Applying this idea to the $1/\sqrt{N}$ problem amounts to overcoming the suppression of the informative part of the total input by postulating that information-carrying signals are correlated or synchronized, perhaps by being locked to various rhythmic patterns of neural activity (Kopell et al., 2000).

Another solution to the $1/\sqrt{N}$ problem is to reduce the effective value of N . In our original estimate, we took N to be the number of synaptic inputs received by a neuron, but it is really only the active synapses that enter into the argument. Sparse coding is a way of reducing N by keeping the number of neurons responding to any given event or feature small. Synapses can be strengthened to compensate for the smaller number of inputs to keep the postsynaptic neuron responding. As discussed above, sparse coding has been proposed as an important principle of neuronal sensory processing for reasons independent of the ones given here. Another way of reducing the effective number of inputs is through local dendritic processing. The idea, advanced most passionately by Bartlett Mel (Poirazi et al., 2003), is that the basic unit of neural computation is not the neuron with its thousands of inputs, but pieces of the dendrite with far fewer. In this case, local dendritic spiking as well as stronger synapses can compensate for the reduced value of N . This work is part of a general revitalization in the theoretical study of dendritic processing (see, for example, Segev and London, 2000) that has been inspired by new data coming from dendritic patching.

Finally, we come to an extremely interesting approach to the problem, which is to suppress the noninformative component of the total input to a neuron through balancing inhibition. Historically, the idea of balancing excitation and inhibition arose

through consideration of the high-degree of variability in neuronal responses (Softky and Koch, 1992), but it can be viewed equally well as a way of canceling positive average inputs. Recall that the noninformative component is the order N contribution coming from the average firing rate of all the presynaptic inputs to a neuron. If we arrange these inputs so that their excitatory and inhibitory effects cancel, it would appear we could eliminate the effects of this component entirely, making the ratio of informative to noninformative signals as large as we want. However, this form of cancellation is not that easy. It is unreasonable to assume that excitatory and inhibitory presynaptic action potentials could be matched perfectly, one-for-one. This means that the subtraction of excitatory input by inhibition is necessarily a noisy business, and the ratio can only be raised to something of order 1. Networks that realize this idea have extremely interesting properties (van Vreeswijk and Sompolinsky, 1996; Brunel, 2000) and provide us with reasonable models of the ongoing, background activity seen in most circuits even in the absence of sensory input or behavioral output. In particular, it seems likely that cortical circuits operate in this “balanced” mode (Shadlen and Newsome, 1994; Troyer and Miller, 1997), and this realization represents an important advance in our understanding of how these circuits operate.

What Are Neural Circuits Doing and How Are They Doing It?

Before we can address how a neural circuit works, we must figure out what it does. Of course, these circuits are responsible for perception, cognition, and action, but the precise roles they play in these processes can be mysterious. Neural models can be divided into two classes depending on whether they address the what or the how of neural responses. Mathematical descriptions of the relationship between neural responses and the stimuli that evoke them can produce precise statements about what a system is doing. Circuit-level models that try to account for these relationships on the basis of known physiological features explore how a circuit works. Of course, many models, including some of those assigned below to one category or another, bridge this distinction. It is nevertheless a useful way to catalog a large body of research.

Except in cases of synaptic transmission by graded potentials, neurons convey information through the timing of their action potentials. (As an aside, a lot of pointless debate over the past 20 years could have been avoided if we had all kept in mind that this statement is *not* necessarily in conflict with the idea that action potential firing can be described by a firing rate.) The work of Bill Bialek and his collaborators has advanced our understanding of and ability to quantify the encoding of information by action potentials through the use of information theory (Rieke et al., 1996). In addition, this research program has spawned new ways of relating neuronal responses to the stimuli that generate them by extended older methods of spike-triggered averaging to spike-triggered covariance (see Rieke et al., 1996) and maximum informative dimensions (Sharpee et al., 2004).

LNP models—which apply a linear filter to a stimulus (L), pass the result of this filtering through a nonlinear function (N), and then generate action potentials by a Poisson process (P)—provide some of our best descriptions of what neurons do. Various

modifications over the past 20 years, including the idea of response normalization (Heeger, 1992) and the realization that optimal solutions can be derived in a wider class of models that include postspike refractoriness and correlations between neurons (Paninski et al., 2007) have broadened the applicability of this approach. These descriptions often include clues about what the underlying biophysical mechanisms might be, so they provide a solid basis for constructing circuit-based models of how responses are generated.

Circuit-level models address how specific patterns of neural activity are generated by sensory stimuli or as part of motor actions. In this context, the theorist's task is to construct a model that reproduces, as much as possible, neural activity recorded during experiments. This form of modeling has been applied extensively to the primary visual cortex where the features to be explained are typically the response selectivities of simple and complex cells (Worgotter and Koch, 1991; Somers et al., 1995; Sompolinsky and Shapley, 1997; Chance et al., 1999; Miller, 2003; Tao et al., 2004). We now have plausible mechanisms for how simple and complex cells obtain their basic response characteristics. Although no single consensus about how the circuits of primary visual cortex operate has arisen from this body of work, this may simply reflect the fact that multiple mechanisms contribute. In other words, many of these ideas are probably correct in one way or another, and the wealth of ideas in this field should be viewed as a success. Circuit-level modeling is now advancing beyond primary sensory areas (for example, Cadieu et al., 2007) and to the consideration of phenomena such as working memory through sustained activity (Amit and Brunel, 1997; Compte et al., 2000; Seung et al., 2000) and decision making (Wang, 2002; Machens et al., 2005).

Physiology and anatomy provide the constraining framework for model building, and we are on the threshold of an explosion in the amount of data available for this purpose due to a major initiative in neuroscience to determine the "connectome." This will involve mapping neural circuits at an unprecedented level of resolution and detail. Theoretical work by Mitya Chklovskii and his collaborators (for example, Wen and Chklovskii, 2008) has helped provide a rationale for some of the basic features of neuroanatomy, but more needs to be done to prepare us for the challenges of this data onslaught. In this regard, it is instructive to consider artificial networks as examples of large complex systems performing useful tasks for which we have complete connectomes. What can we learn from the complete connectome or, indeed, a complete mathematical description of a complex artificial network model?

First, what can't we learn? It is unlikely, for example, that we could deduce the task that the network was constructed to perform even if we were given the complete equations and connections of the model. If, along with this information, we were told what this task was, it is unlikely that we could figure out how the network performs it. If we somehow managed to make any progress along these lines, the people who constructed the network could probably provide us with another one that performs the same task but has a different connectome. In a similar way, biological systems may operate in a more variable manner than we have suspected, as has been stressed by Eve Marder (Marder et al., 2007). These issues are particularly true of a class

of network models known as liquid state or echostate networks (Maass et al., 2002; Jaeger, 2003). In these models, the vast majority of interneuronal connections are not directly related to the task being performed (they are typically chosen randomly and left unchanged), the exceptions being synapses onto the output units of the network. Nevertheless, the tuned values of the synapses onto the output units can only be understood through their relationships to the random synapses. Such systems represent enormous challenges for conventional anatomical and physiological approaches.

The fact that the connectome of an artificial neural network does not typically tell us what the network does or how it does it should not be taken as an indication that this information is useless. Far from it. But we must be willing to be more abstract in our thinking. The important issue for an artificial network is not how it works but how it was constructed, which means what training procedures and modification rules were used to get it to perform a task. Although this information is not provided directly by the connectome, much can be inferred. For example, it is important to know whether the network has a feedforward architecture or has strong feedback loops. Other features of the network layout, whether it has hubs or bottlenecks, how many layers it contains, and its degree of heterogeneity, provide important clues as well. Obtaining a high-resolution connectome in neuroscience will be of great value, but artificial neural networks provide a cautionary tale that reminds us that scientific revolutions tend to render uninteresting as many questions as they answer. We will be fortunate if the connectome project does this for neuroscience, but as we launch ourselves into it we should appreciate that, as artificial neural networks appear to suggest, we may be asking the wrong questions.

What Matters and What Doesn't?

Identifying the minimum set of features needed to account for a particular phenomenon and describing these accurately enough to do the job is a key component of model building. Anything more than this minimum set makes the model harder to understand and more difficult to evaluate. The term "realistic" model is a sociological rather than a scientific term. The truly realistic model is as impossible and useless a concept as Borges' "map of the empire that was of the same scale as the empire and that coincided with it point for point" (Borges, 1975). In any model, a phenomenon must be accounted for by an approximate description of a subset of the features that exist in the natural system. The art of modeling lies in deciding what this subset should be and how it should be described. Testing the implications of including different features is part of the ongoing activity that theoretical neuroscientists need to do to advance the field.

An interesting lesson concerning approximation in modeling arises from the mathematical and computational work on synaptic mechanisms of memory storage mentioned in the introduction. This research involved considerable simplifications of the relevant biology but, nevertheless, the message it generated—that neural networks could store and recall huge amounts of information through synaptic plasticity—seemed so clear and unequivocal that it appeared to transcend these simplifications. It wasn't until considerably later that Stefano Fusi and Daniel Amit revealed a fly in the ointment (Amit and Fusi, 1994). The

bulk of work on memory storage involved constructing networks that retained information about pre-existing lists of “memories.” Biological memory storage requires a continuous influx of information to be stored, and this introduces severe problems because new memories tend to overwrite old ones (Grossberg, 1982). Amit and Fusi (1994) showed that one of the approximations in the original modeling work, allowing synaptic strengths to vary without bound, has enormous implications when a network operates in such a continuous manner. Eliminating this approximation catastrophically reduces the memory storage capacity of the models. The unbounded nature of the synapses was not an approximation that introduced a modest error into the results, but rather something that entirely changed the way the network worked. These are the kinds of approximations we must be careful to avoid. The implications of this discovery, which are still being worked out, are likely to revise our picture of memory storage through synaptic plasticity dramatically (Fusi, 2002; Fusi et al., 2005).

How are we to find our way through the morass of complexity we face in even the simplest of neural systems? Describing the full range of this complexity is impossible. Acknowledging that the system is complex and heterogeneous may be more important than accounting for exactly how it is complex. This does not, however, represent less of a modeling challenge. Our models may not have to be accurate in their details, but they must capture the wide range of the phenomena operating in neural circuitry. In other words, breadth is more important than depth in neural modeling.

The Future

Neural modelers tend to divide problems into fast and slow components and, in most applications, the fast part is performing a particular task whereas the slow part is learning it. The enormous amount of attention given to long-term synaptic plasticity illustrates that learning is widely considered a job for the synapse. Another aspect of this slower process that has received both modeling and experimental attention is the modulatory reinforcement needed to guide learning (Schultz et al., 1997). However, once established by a plasticity procedure, synapses in most models act as passive conveyors of information during the faster performance of the learned task. A notable exception to this is network models in which short-term plasticity plays an important functional role, as typified by the work of Misha Tsodyks (most recently in Mongillo et al., 2008). Nevertheless, we more commonly tend to think of synapses as the locus of learning and memory, and neurons as the workhorses of dynamic computation. This may be radically wrong. Synapses are extraordinarily complex biochemical devices that are likely to have computational potential far beyond what we currently model or understand (Emes et al., 2008). We may encounter revelatory surprises as these capabilities are uncovered experimentally and explored theoretically.

As is often noted, computers can outperform humans by many orders of magnitude in arithmetic but are rather pathetic at visual and auditory recognition tasks that we find effortless. We may outperform machines in these tasks by being extraordinarily skilled and prolific guessers and by being remarkably quick at using small amounts of evidence to eliminate wrong guesses

or confirm correct ones. Three elements are at play here: instant access to very large memory stores, the ability to generate hypotheses, and an interaction between internally generated hypotheses and external evidence that allows sensory data to veto or support internal constructs extremely efficiently. Google appears to have solved the first of these requirements in a machine setting, but the latter two are currently beyond our understanding, either in the world of computers or in neuroscience (although generative models provide a start). This is where I think the future lies in theoretical investigations of cognitive function. We must learn how to build models that construct hypotheses through their internally generated activity while remaining sensitive to the constraints provided by externally generated sensory evidence. As outlined in this review, the past 20 years of research has produced interesting models of internally generated activity and of sensory-driven responses, but the marriage of these two components of cognition remains a challenge for the future.

ACKNOWLEDGMENTS

I thank Stefano Fusi for valuable comments and suggestions and Sara Solla for pointing out the relevance of the Borges story (Borges, 1975).

REFERENCES

- Amit, D., and Fusi, S. (1994). Dynamic learning in neural networks with material synapses. *Neural Comput.* 6, 957–982.
- Amit, D.J., and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* 7, 237–252.
- Amit, D.J., Gutfreund, H., and Sompolinsky, H. (1985). Spin-glass models of neural networks. *Phys. Rev. A* 32, 1007–1018.
- Atick, J.J., and Redlich, A.N. (1990). Towards a theory of early visual processing. *Neural Comput.* 2, 308–320.
- Barlow, H. (2001). Redundancy reduction revisited. *Network* 12, 241–253.
- Bell, A.J., and Sejnowski, T.J. (1997). The ‘independent components’ of natural scenes are edge filters. *Vision Res.* 37, 3327–3338.
- Borges, J.L. (1975). On exactitude in science. In *A Universal History of Infamy* (London: Penguin).
- Brunel, N. (2000). Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. *J. Physiol. (Paris)* 94, 445–463.
- Buchs, N.J., and Senn, W. (2002). Spike-based synaptic plasticity and the emergence of direction selective simple cells: simulation results. *J. Comput. Neurosci.* 13, 167–186.
- Cadieu, C., Kouh, M., Pasupathy, A., Connor, C.E., Reisenhuber, M., and Poggio, T. (2007). A model of V4 shape selectivity and invariance. *J. Neurophysiol.* 98, 1733–1750.
- Chance, F.S., Nelson, S.B., and Abbott, L.F. (1999). Complex cells as cortically amplified simple cells. *Nat. Neurosci.* 2, 277–282.
- Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb. Cortex* 10, 910–923.
- Dan, Y., and Poo, M.M. (2006). Spike timing-dependent plasticity: from synapse to perception. *Physiol. Rev.* 86, 1033–1048.
- Dong, D.W., and Atick, J.J. (1995). Temporal decorrelation: a theory of lagged and non-lagged responses in the lateral geniculate nucleus. *Network Comput. Neural Syst.* 6, 159–178.
- Emes, R.D., Pocklington, A.J., Anderson, C.N., Bayes, A., Collins, M.O., Vickers, C.A., Croning, M.D., Malik, B.R., Chourdary, J.S., Armstrong, J.D., and

- Grant, S.G. (2008). Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* *11*, 799–806.
- Ernst, M.O., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* *415*, 429–433.
- Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* *4*, 2379–2394.
- Fusi, S. (2002). Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. *Biol. Cybern.* *87*, 459–470.
- Fusi, S., Drew, P., and Abbott, L.F. (2005). Cascade models of synaptically stored memories. *Neuron* *45*, 599–611.
- Gerstner, W., Kemper, R., van Hemen, J.L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* *383*, 76–81.
- Grossberg, S. (1982). Processing of expected and unexpected events during conditioning and attention: a psychophysiological theory. *Psychol. Rev.* *89*, 529–572.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent selective computational abilities. *Proc. Natl. Acad. Sci. USA* *79*, 2554–2558.
- Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* *9*, 181–198.
- Hinton, G.E., and Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* *352*, 1177–1190.
- Hinton, G.E., Osindero, S., and Teh, Y.W. (2006). A fast learning algorithm for deep belief nets. *Neural Comput.* *18*, 1527–1554.
- Izhikevich, E.M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* *17*, 2443–2452.
- Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks. In *Advances in Neural Information Processing Systems*, *15*, S. Becker, S. Thrun, and K. Obermayer, eds. (Cambridge, MA: MIT Press), pp. 593–600.
- Kang, S., Kitano, K., and Fukui, T. (2008). Structure of spontaneous UP and DOWN transitions self-organizing in a cortical network model. *PLoS Comput. Biol.* *4*, e1000022.
- Knill, D.C., and Richards, W. (1996). *Perception as Bayesian Inference* (New York: Cambridge University Press).
- Kopell, N., Ermentrout, G.B., Whittington, M.A., and Traub, R.D. (2000). Gamma rhythms and beta rhythms have different synchronization properties. *Proc. Natl. Acad. Sci. USA* *97*, 1867–1872.
- Kording, K.P., and Wolpert, D.M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* *10*, 319–326.
- Lewicki, M.S. (2002). Efficient coding of natural sounds. *Nat. Neurosci.* *5*, 356–363.
- Ma, W.J., Beck, J.M., Latham, P.E., and Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nat. Neurosci.* *9*, 1432–1438.
- Maass, W., Matsuura, T., and Markram, H. (2002). Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* *14*, 2531–2560.
- Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* *30*, 1121–1124.
- Marder, E., Tobin, A.E., and Grashow, R. (2007). How tightly tuned are network parameters? Insight from computational and experimental studies in small rhythmic motor networks. *Prog. Brain Res.* *165*, 193–200.
- Miller, K.D. (2003). Understanding layer 4 of the cortical circuit: A model based on Cat V1. *Cereb. Cortex* *13*, 73–82.
- Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. *Science* *319*, 1543–1546.
- Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Comput.* *19*, 1437–1467.
- Niven, J.E., Anderson, J.C., and Laughlin, S.B. (2007). Fly photoreceptors demonstrate energy-information trade-offs in neural coding. *PLoS Biol.* *5*, e116.
- Nowotny, T., Zhigulin, V.P., Selverston, A.I., Abarbanel, H.D., and Rabinovich, M.I. (2003). Enhancement of synchronization in a hybrid neural circuit by spike-timing dependent plasticity. *J. Neurosci.* *23*, 9776–9785.
- Olshausen, B.A., and Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* *381*, 607–609.
- Paninski, L., Pillow, J., and Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Prog. Brain Res.* *165*, 493–507.
- Poirazi, P., Brannon, T., and Mel, B.W. (2003). Pyramidal neuron as 2-layer neural network. *Neuron* *37*, 989–999.
- Rao R.P.N., Olshausen B.A., and Lewicki M.S., eds. (2002). *Probabilistic Models of the Brain: Perception and Neural Function* (Cambridge, MA: MIT Press).
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1996). *Spikes: Exploring the Neural Code* (Cambridge, MA: MIT Press).
- Ruderman, D.L., and Bialek, W. (1994). Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.* *73*, 814–817.
- Rumelhart D.E. and McClelland J.L., eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1, Foundations* (Cambridge, MA: MIT Press).
- Salinas, E., and Sejnowski, T.J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* *2*, 539–550.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. *Science* *275*, 1593–1599.
- Segev, I., and London, M. (2000). Untangling dendrites with quantitative models. *Science* *290*, 744–750.
- Seung, H.S., Lee, D.D., Reis, B.Y., and Tank, D.W. (2000). Stability of the memory for eye position in a recurrent network of conductance-based model neurons. *Neuron* *26*, 259–271.
- Shadlen, M.N., and Newsome, W.T. (1994). Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* *4*, 569–579.
- Sharpee, T., Rust, N.C., and Bialek, W. (2004). Analyzing neural responses to natural signals: maximally informative dimensions. *Neural Comput.* *16*, 223–250.
- Softky, W.R., and Koch, C. (1992). Cortical cells should spike regularly but do not. *Neural Comput.* *4*, 643–646.
- Somers, D.C., Nelson, S.B., and Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.* *15*, 5448–5465.
- Sompolinsky, H., and Shapley, R. (1997). New perspectives on the mechanisms for orientation selectivity. *Curr. Opin. Neurobiol.* *7*, 514–522.
- Song, S., and Abbott, L.F. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron* *32*, 339–350.
- Stocker, A.A., and Simoncelli, E.P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* *9*, 578–585.
- Tao, L., Shelley, M., McLaughlin, D., and Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proc. Natl. Acad. Sci. USA* *101*, 366–371.
- Tolhurst, D.J., Tadmor, Y., and Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic Physiol. Opt.* *12*, 229–232.
- Troyer, T.W., and Miller, K.D. (1997). Physiological gain leads to high ISI variability in a simple model of a cortical regular spiking cell. *Neural Comput.* *9*, 971–983.
- van der Schaaf, A., and van Hateren, J.H. (1996). Modelling the power spectra of natural images: statistics and information. *Vision Res.* *28*, 2759–2770.
- van Hateren, J.H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. *Vision Res.* *37*, 3407–3416.

van Vreeswijk, C., and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726.

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36, 955–968.

Wen, Q., and Chklovskii, D.B. (2008). A cost-benefit analysis of neuronal morphology. *J. Neurophysiol.* 99, 2320–2328.

Womelsdorf, T., and Fries, P. (2007). The role of neuronal synchronization in selective attention. *Curr. Opin. Neurobiol.* 17, 154–160.

Worgotter, F., and Koch, C. (1991). A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *J. Neurosci.* 11, 1959–1979.