Primer

Recurrent neuronal circuits in the neocortex

Rodney J. Douglas and Kevan A.C. Martin

The neocortex is a remarkable computational system. It uses slow computational elements to solve in real time a variety of computational problems that cannot be achieved by man-made computational methods implemented on very much faster hardware. This performance is clearly due to the fundamentally different computational methods that biology has evolved under very tight constraints of energy consumption and necessity for speed. Our brain consumes approximately 20W whereas a silicon version of our brain built with present-day chip technology would consume 10MW — it would melt.

While understanding the coding in nerve impulses is generally regarded as the key to understanding the brain's computations, it is clear that the vast majority of computations in the brain are done by slowly varying analogue potentials in the richly branched dendrites of neurons. How is it then that a human brain, such as that of Garry Kasparov, can be competitive with machines like Deep Blue, the chess-playing computer that evaluates 200 million possible moves per second? Understanding biology's methods of computation would open the way to explaining such human abilities, as well as allow the development of novel processing technologies. For these reasons there has been a sharp increase in research on the architecture of neocortical circuits, since they provide the physical support for neocortical processing. In this Primer, we shall describe one interesting property of neocortical circuits — recurrent connectivity — and suggest what its computational significance might be.

Most cortical connections are local and excitatory

Although the human cortex contains about 10^9 neurons, the basic architecture of the cortex can be understood in terms of the laminar distribution of relatively few types (say about 100) of excitatory and inhibitory neurons. The degree of connection between these types of neuron has been estimated for the static anatomy of cat visual cortex (Figure 1) and also for physiological connections between some neuronal types in rat somatic cortex. These and other quantitative anatomical circuit data suggest a number of intriguing circuit properties, which offer many intriguing clues to the fundamental properties of cortical processing.

One clue is the dominance of local cortical synapses over those provided by individual afferents of a given cortical area, as shown in Figure 1. Overall, the vast majority of cortical excitatory synapses and virtually all inhibitory synapses originate from neurons within cortex and most of these synapses originate from neurons within the local cortical area. This means that the afferent projections of cortex, from the thalamus and other individual cortical areas, each comprise a surprisingly small percentage of all excitatory synapses in the target area. For example, in the visual cortex, synapses from the lateral geniculate nucleus of the thalamus form only 5–10% of the excitatory synapses, even in their main target layer (layer 4) of area 17 in cats and monkeys, yet these afferents clearly provide sufficient excitation to drive the cortex.

Similarly, the projections that connect different areas of neocortex also form a fraction of a percent of the synapses in their target layers, yet both the 'feedforward' and 'feedback' inter-areal circuits are functionally significant. This raises the critical question of how the local cortical circuits reliably process the seemingly small input signals that arise from peripheral sense organs, or within the cortex itself, and how it is that the fidelity of these signals is retained as they are transmitted through the hierarchy of cortical areas?

A second clue is the large fraction of excitatory connections made between pyramidal cells of the superficial cortical layers (Figure 1). Although excitatory neurons in other layers, such as the spiny stellate neurons of layer 4, also receive input from their neighbours, it is only in the superficial layers that the

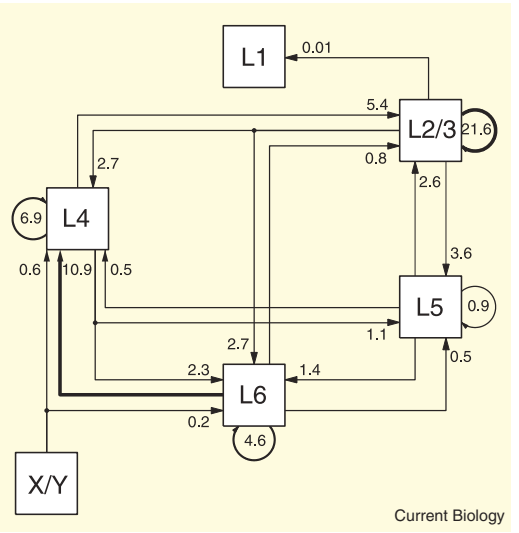

Figure 1. A quantitative graph of the connections between various classes of excitatory neurons and their targets in cortex.
Only the connections between the classes of the dominant excitatory cell types are shown in this partial diagram. Each arrow is labeled with a number indicating the proportion of all the excitatory synapses in area 17 that are formed between the various classes of excitatory neurons. Total number of synapses between excitatory neurons is 13.6×10^{10}. Additional maps of connections from excitatory to inhibitory neurons, and so on, can be found in Binzegger *et al.* (2004).

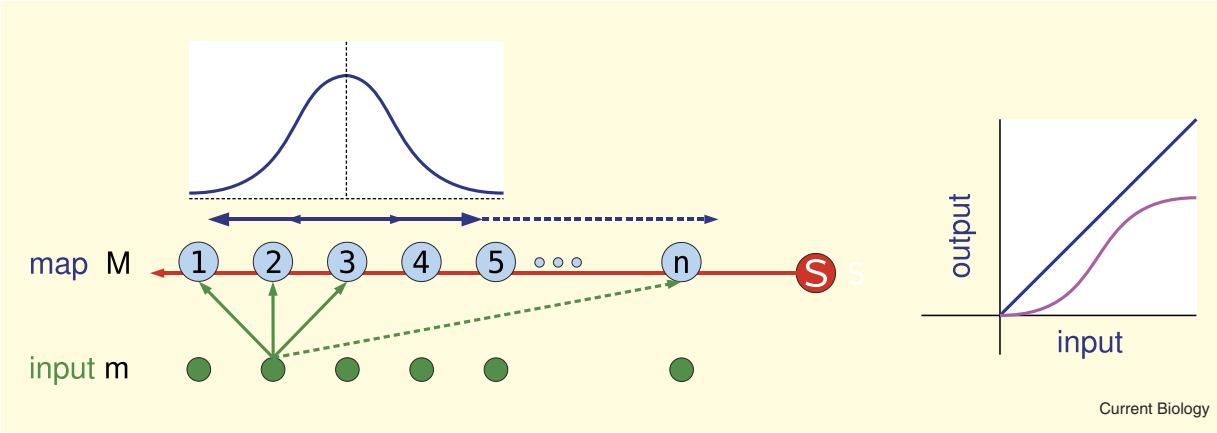

Figure 2. Simple model of recurrently connected neurons with rate coded outputs.

A population of N rate excitatory neurons (blue filled circles) arranged as a spatial array; each receives inhibition from a common inhibitory neuron (red filled circle), and each of them excites that inhibitory neuron. Each excitatory neuron receives feedforward excitatory input (green arrows from green input neurons), as well as recurrent excitatory input from their close neighbours. The strength of the recurrent inputs made onto any target neuron is a bell shaped function of the neighbour's displacement in the array from the target neuron (example shown for neuron 3). The activation function (right) of all neurons is a thresholded linear function (blue line) rather than a sigmoid (cyan). Unlike the sigmoid, the linear activation is unbounded above so that the stability of the network must depend on the integration of excitatory and inhibitory neurons rather than the sigmoidal saturation of individual neurons.

pyramidal cells make very extensive arborizations within their same layer. Indeed, nearly 70% of a superficial pyramidal cell's excitatory input is derived from other cells of its own type. Consequently, first-order recurrent connections between layer 2 and 3 pyramidal cells, in which a target neuron projects back to its source neuron in a tight positive feedback loop, are more likely than in any other layer. We have proposed that positive feedback plays a crucial role in cortical computation by providing gain for active selection and re-combination of the relatively small afferent signals.

Control of gain is a critical aspect of cortical computation

The gain of a system is the (dimensionless) ratio of the magnitudes of two causally related signals. In a feedback system, two different gains are usually considered. The first is the overall 'system gain'. This is measured as a ratio of the output over the input of the system. The second is the 'loop gain'. This is measured around the feedback loop, and can be expressed as the fraction of the output signal that is due to feedback. Thus, when the loop gain is zero, the system gain is entirely feedforward. As the loop gain approaches one, the system gain becomes dominated by its feedback. If the loop gain exceeds one, the system is unstable and its output diverges.

The feedforward gain of individual neurons operating in rate mode is small. Typically, many input spike events must be applied to a neuron before it produces a single spike output. However, simulation studies and physiological evidence suggest that cortical circuits can generate significant system gain by the positive feedback excitation mediated by recurrent intracortical axonal connections (Figure 2). Positive feedback amplification may seem inherently dangerous, but neurons subject to positive feedback can be stable if the sum of the excitatory currents evoked by the afferent input and the positive feedback excitatory currents is less than the total negative current dissipated though their membrane leak conductances, the active conductances of the action potential, and active membrane and synaptic conductances.

Such circuit-dependent stability must exist in cortex, because the steady discharge rate of active cortical neurons is usually much less than their maximum rate, so stability does not depend on the neurons being driven into discharge saturation. It is this positive feedback amplification that allows a small input signal to be 'heard' in cortex. The question, of course, is how is this small signal ever distinguished from 'spontaneous' cortical activity. Here simplified artificial network models have been invaluable in providing insights into the properties of recurrent networks.

Recurrent circuits perform signal restoration (and much besides)

Although artificial network models are much simpler than networks of real cortical neurons, they have the advantage that their modes of behaviour can be clearly understood and then used to interpret the experimentally observed organization and operation of cortical networks. For example, Hopfield and others showed that recurrent networks of ideal neurons are dynamical systems whose stable patterns of activation (or attractors) can be viewed as memories, or as the solutions to constraint satisfaction problems. More recently, there has been a growth of interest in the use of networks of 'linear threshold

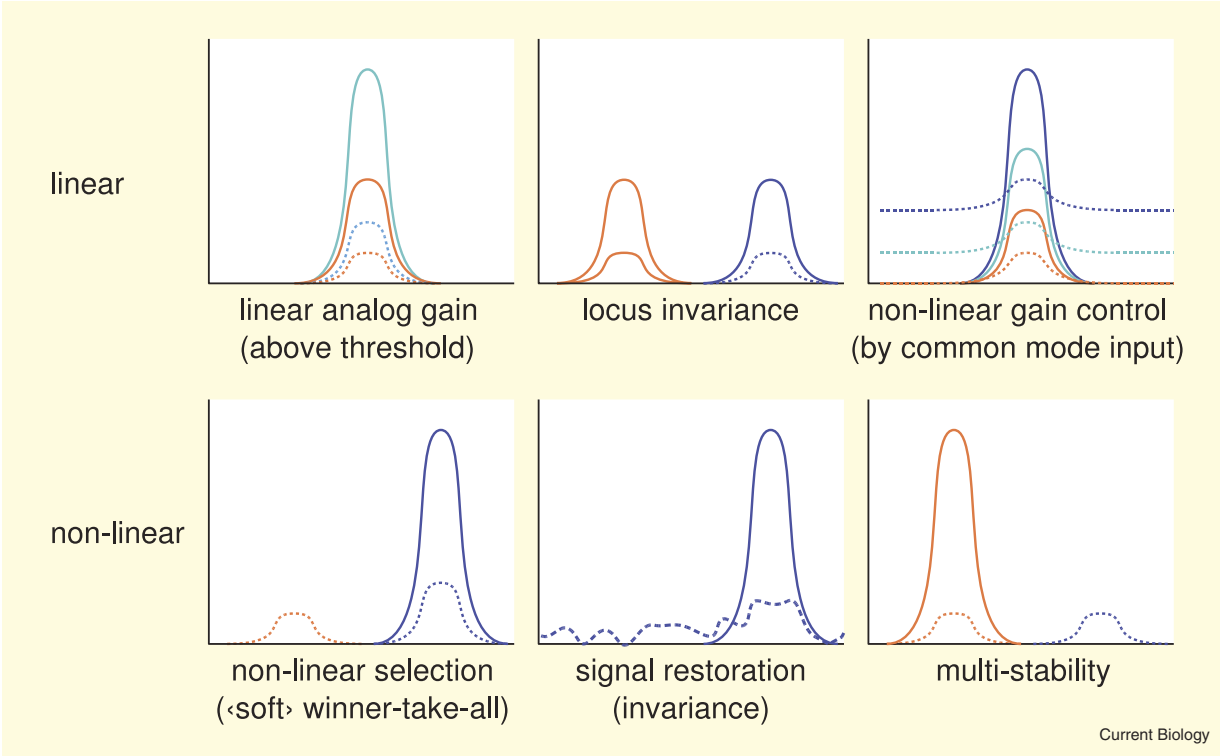

Figure 3. Six interesting functional properties of the recurrent network described in Figure 2.

In this multi-part figure, each part shows the response of the array of excitatory neurons (along the x-axis). Top left: Linear Gain. Above threshold, the network amplifies its hill-shaped input (stipled lines) with constant gain (output, solid lines). Top center: Locus Invariance. This gain is locus invariant (provided that the connections' weights are homogenous across the array). Top right: Gain Modulation. The gain of the network can be modulated by an additional constant input applied to all the excitatory neurons, and superimposed on the hill-shaped input. The gain is least when no constant input is applied (input, red stippled line; output, red solid line), and largest for a large constant input (blue lines). Bottom left: Winner-take-all. When two inputs of different amplitude are applied to the network, it selects the stronger one. Bottom center: Signal Restoration. The network is able to restore the hill-shaped input, even when that input is embedded in noise. Bottom right: Bistability. When separate inputs have the same amplitudes, the network selects one, according to its initial conditions at the time the input is applied.

neurons' (LTNs) to understand cortical circuits. LTNs have continuous valued (non-spiking) positive outputs that are directly proportional to the positive difference between the excitation and inhibition that they receive. If this difference is zero or negative, they remain silent. LTN neurons are interesting because their threshold behaviour and linear response properties resemble those of cortical neurons.

A commonly studied LTN network consists of two populations of neurons, as illustrated in Figure 2. One population consists of excitatory neurons and a smaller population of inhibitory neurons (even one global inhibitory neuron will suffice, as in Figure 2). For simplicity, the patterns of connection within each population are homogenous. The excitatory neurons receive feedforward excitatory connections that carry the input signal, feedback excitatory connections from other members of their population, and feedback inhibitory connections from the inhibitory neuron(s). Often the populations of excitatory neurons are arranged as a one-dimensional spatial map and the pattern of their recurrent connection strengths is regular, which typically is expressed as a hill-shaped function of distance of a source neuron from its target.

Even such simple recurrent networks have interesting properties that contribute directly to our understanding of signal processing by the neuronal circuits of cortex. The properties illustrated in Figure 3 arise out of the interaction between the feedback excitation, which amplifies the inputs to the network, and the non-linearity introduced by the inhibitory threshold, which itself depends on the overall network activity. The important result to note here is that the positive feedback enhances the features of the input that match patterns embedded in the weights of the excitatory feedback connections, while the overall strength of the excitatory response is used to suppress outliers via the dynamical inhibitory threshold imposed by the global inhibitory neurons. In this sense, the network can actively impose an interpretation on an incomplete or noisy input signal by restoring it towards some fundamental activity distribution embedded in its excitatory connections — the cortical 'hypothesis'.

The explanation for this remarkable processing property

is as follows. Consider the network illustrated in Figure 2. A weight matrix can describe the synaptic interactions between the various neurons of the network. Notice, however, that if a neuron is not active, it is effectively decoupled from the circuit and so does not express its interactions. In this case the full weight matrix of the network can be replaced by a reduced matrix, the 'effective weight matrix', which is similar to the full matrix, but with all entries of the silent neurons zeroed out. Consequently, as various neurons rise and fall across their discharge threshold, the effective weight matrix changes. Some of these matrices may be stable, but others may not be so. We will consider how the network can converge to its steady-state output by passing through various combinations of active neurons and thus various effective weight matrices.

Imagine that a constant input pattern is applied to the input neurons. If, in the unlikely case that the outputs do not change at all, then the job is done: the network has already converged. More likely, some or all of the neurons will change their activity as a result of a combination of feedforward and feedback activation. One possibility is that all the neurons could increase their activity as a result of unstable positive feedback. But this instability is forbidden by the common inhibition applied to all excitatory neurons, which, provided the inhibition is strong enough, precludes a regenerative common increase in activation. The remaining possibility is that although the feedback is unstable, only some neurons are able to increase their activation, while others must decrease. Then, eventually one of the decreasing neurons will fall beneath threshold, at which stage it no longer contributes to the active circuit and so the effective weight matrix must change, removing the interactions of this newly silent neuron. This pruning process continues until finally the network selects a combination of neurons (a 'permitted set'), the effective weight matrix of which is stable, and allows the network to converge to a steady state. The important observation here is that positive feedback can be unstable — the feedback gain is greater than one — during the transient behavior of the networks, and that the network can use this instability to explore new partitions of active neurons until a suitable (stable) partition, consistent with the input pattern, is found. It is in this modulation of the strength of positive feedback that the computationally interesting properties of the recurrent cortical circuits rests.

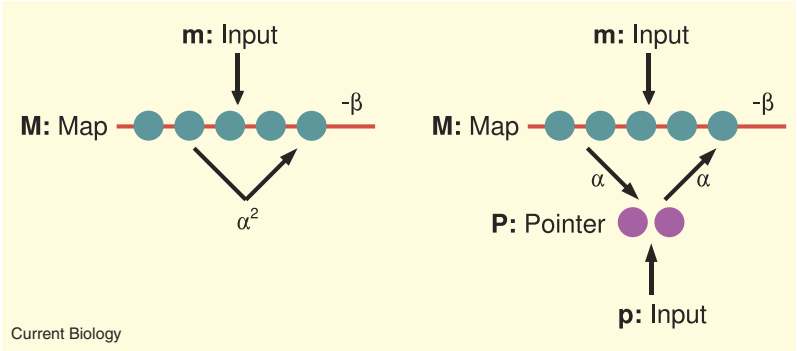

Figure 4. Comparison of standard recurrently connected network left (equivalent to Figure 2), and the 'pointer-map' configuration of recurrence.

Both networks have inhibitory feedback (–β, red). In both cases the overall feedback between excitatory neurons is the same. In the standard network this feedback is applied monosynaptically. In the pointer-map, a small, for example two cell, population of 'pointer neurons' is inserted in the feedback loop. In this way, the feedback is decomposed into two successive stages, each providing gain α. The two pointer neurons have differently biased connectivities to the map of excitatory neurons. The left pointer neuron is more strongly connected to the leftmost neurons of the map, and the right pointer to the rightmost neurons of the map. If the pointer neurons are not perturbed by their inputs (p) then the pointer-map behaves like the simple recurrent network at left. When the 'feedback' or 'top down' input p is applied, it differentially activates the pointer neurons, and so biases the distribution of feedback gain to the map. For example, if input p is applied only to the left pointer, amplification of 'bottom up' map input m will be increased towards the left of the map, and reduced toward the right, so providing an attentional focus toward the left.

How 'top-down' connections steer local circuits

The control of gain within the local recurrent circuits illustrates how patterns of thalamic input can be effectively selected. This is possible because the thalamic inputs map topographically across a cortical area and because the recurrent circuits are local. Similar considerations probably apply to the processing of 'feedforward' inputs from other cortical areas, which, like the thalamic input, target the middle layers of the cortex. However, 'feedback' connections, which target the superficial and deep layers of cortex, are usually thought to be too weak to drive the local circuit. Nonetheless, it is clear that the response of cortical neurons to an appropriate stimulus changes dramatically according to whether a stimulus is being attended to or not; these changes are thought to be mediated by feedback or 'top down' projections. These modulations in the magnitude of a response can be interpreted as changes in the gain of the local circuit controlled by an external attentional input, for example from another cortical area.

What sort of circuit could achieve this? Figure 4 illustrates one simple idea, which is an extension of the concept of the 'permitted set'. In this network, the recurrent activity is not transmitted monosynaptically between members of the cortical map, but instead via a group of 'pointer neurons'. The pointer neurons, which form only a small

fraction of the total neurons in the map, distribute their output, or 'point', to different regions of the cortical map. The strength of their pointing is determined by the external inputs, which then have control over the gain of the local circuits that form the cortical map. Thus in a situation where there are two competing inputs supplied by the feedforward circuits, the activity of one can be enhanced at the cost of the other by the pointer neuron system.

Conclusions

A surprising, but consistent pattern across all areas of neocortex examined, is that the most of the thousands of synapses formed on the dendritic tree of any neuron come from its neighbouring excitatory neurons located in the same cortical area. Very few synapses are contributed by long distance connections, whether they arise from neurons in subcortical nuclei (principally the thalamus) or other cortical areas. Thus, the local circuit is the heart of cortical computation. While limitations on connectivity and power, and the necessity of robustness and reliability, have no doubt propelled the evolution of this kind of neocortical architecture, the consequences of this architecture have been difficult to study experimentally, particularly so in the case of the physiological interactions of local versus long distance projections. Thus, although we have developed very powerful tools, such as molecular and optical methods, for observing fine details of the structure and function of neocortical circuits, the challenge of understanding how these circuits actually work and what they actually do has not diminished. Here we have offered an interpretation of the neocortical architecture and have used simple models to illustrate some possible mechanisms by which the local recurrent circuits behave and interact with the long distance 'feedforward' or 'feedback' projections. Already these rather simple circuits provide a rich set of behaviours that are consistent with known computations of the neocortex.

Acknowledgments

We thank our colleagues John Anderson and Tom Binzegger for their collaboration; and EU grant DAISY (FP6-2005-015803) for financial support.

Further reading

Anderson, J.C., Binzegger, T., Martin, K.A.C., and Rockland, K.S. (1998). The connection from area V1 to V5: A light and electron microscopic study. J. Neurosci. *18*, 10525–10540.

Anderson, J.C., and Martin, K.A.C. (2006). Synaptic connection from cortical area V4 to V2 in macaque monkey. J. Comp. Neurol. *495*, 709–721.

Ben-Yishai, R., Bar-Or, R.L. and Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. Proc. Natl. Acad. Sci. USA *92*, 3844–3848.

Binzegger, T., Douglas, R.J., and Martin, K. A.C. (2004). A quantitative map of the circuit of cat primary visual cortex. J. Neurosci. *24*, 8441–8453.

Braitenberg, V., and Schüz, A. (1998). Cortex: Statistics and Geometry of Neuronal Connections, Second Edition (Springer-Verlag: Heidelberg).

Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A., and Suarez, H.H. (1995). Recurrent excitation in neocortical circuits. Science *269*, 981–985.

Douglas, R.J. and Martin, K.A. (1991). A functional microcircuit for cat visual cortex. J. Physiol. *440*, 735–769.

Douglas, R.J. and Martin, K.A.C. (2004). Neuronal circuits of the neocortex. Annu. Rev. Neurosci. *27*, 419–451.

Hahnloser, R. H., Sarpeshkar, R., Mahowald, M.A., Douglas, R.J., and Seung, H.S. (2000). Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature *405*, 947–951.

Hahnloser, R.H.R., Seung, H.S., and Slotine, J.-J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. Neural Comput. *15*, 621–638.

Hansel, D. and Sompolinsky, H. (1998). Modeling feature selectivity in local cortical circuits. In Methods in Neuronal Modeling: From Synapse to Networks, C. Koch and I. Segev eds., Second Edition (Cambridge, MA: MIT Press), pp. 499–567.

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. USA *79*, 2554–2558.

Salinas, E., and Abbott, L.F. (1996). A model of multiplicative neural responses in parietal cortex. Proc. Natl. Acad. Sci. USA *93*, 11956–11961.

Shipp, S. (2007). Structure and function of the cerebral cortex. Curr. Biol. *17*, R1–R6.

Thomson, A.M., West, D.C., Wang, Y., and Bannister, A.P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2-5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. Cereb. Cortex *12*, 936–953.

Institute of Neuroinformatics, Winterthurerstrasse 190, 8057 Zurich, Switzerland. E-mail: kevan@ini.phys.ethz.ch

Book review

Diseases desperate grown

Walter Gratzer

The Cancer Treatment Revolution – How Smart Drugs and Other Therapies Are Renewing Our Hope and Changing the Face of Medicine David G. NathanWiley, New Jersey, 2007
ISBN 978-0-471-94654-0

"We all labour against our own cure; for death is the cure of all diseases." So reflected Sir Thomas Browne, the great 17th-century physician and aphorist. In our society today we labour even harder; death is no longer seen as a valid option, more as an affront to medical science. Well within living memory cancer treatment was a hit-and-miss affair. The way George Bernard Shaw saw it, if the patient died it was put down to 'natural causes', while if he lived, the doctor took the credit. The available treatments were surgery, radiation and a little later, chemotherapy of an unsparing kind, likened by David Nathan to carpet-bombing — destruction, that is, of the entire terrain in the hope of knocking down the target before the patient. But now such butchery is being edged out, as Nathan's title implies, by the rigours of science, and the skills of the 'physician scientists', virtuosos of the genome and the proteome, no less than of the stethoscope and scalpel. Drugs of a new kind are directed not just at malignant cells, but at those of the particular malignancy, and they have induced recoveries unimaginable a decade or two ago.

David Nathan is an eminent doctor, researcher and teacher, whose academic progeny populate medical faculties around the world. He speaks, we may take it, with authority, and this book now reveals him to be a writer of enviable lucidity and style. He has built his narrative round the fortunes of three patients with