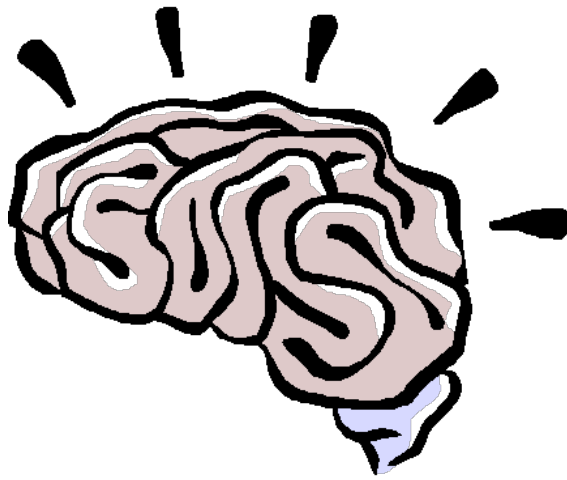


COMPUTATIONAL COGNITIVE NEUROSCIENCE : A READER



NEURONAL VARIABILITY: NOISE OR PART OF THE SIGNAL?

Richard B. Stein^{*‡}, E. Roderich Gossen[§] and Kelvin E. Jones^{‡§}

Abstract | Sensory, motor and cortical neurons fire impulses or spikes at a regular, but slowly declining, rate in response to a constant current stimulus. Yet, the intervals between spikes often vary randomly during behaviour. Is this variation an unavoidable effect of generating spikes by sensory or synaptic processes ('neural noise') or is it an important part of the 'signal' that is transmitted to other neurons? Here, we mainly discuss this question in relation to sensory and motor processes, as the signals are best identified in such systems, although we also touch on central processes.

SPIKE TRAIN

A temporal sequence of all-or-none action potentials.

INTER-SPIKE INTERVAL

The time between two successive spikes in a train.

We can quickly recognize the face of a family member from among the billions of faces on our planet or identify a friend at some distance from their pattern of walking, even in a crowd. The basis for this remarkable ability must reside in the series of impulses (spikes) in individual nerve cells or populations of nerve cells. Neuroscientists have long debated how SPIKE TRAINS code information about sensory events. When Lord Adrian and colleagues first took recordings from sensory and motor neurons more than 75 years ago, they found that the intensity of a stimulus was coded as a rate of nerve impulses over time^{1,2}, which is now referred to as the 'rate code'. As described below, others have questioned whether rate is the only parameter that encodes information about sensory and motor events. Does the exact timing of spikes from one neuron or the relative timing of spikes in a population of neurons convey further information that is important in the feats of pattern recognition that are mentioned above? These concepts have been referred to as 'temporal coding'. This topic has recently assumed new importance because of the implantation of an array of electrodes into the motor cortex of a severely disabled individual³. The spikes from these neurons are being processed to move a cursor on a computer screen in real time. The best method for decoding these spike trains will depend on our understanding, in detail, how information about movements is normally coded in the motor cortex.

The specific question raised in the title of this review follows on from these general issues. If a steady rate is important to determine the intensity of a constant sensory or motor signal, then any variations in the INTER-SPIKE INTERVALS will cause fluctuations in the rate, which might be considered unwanted noise. However, if the timing of each spike carries extra information, then the variability can be an important part of the signal. We argue here that both temporal and rate coding are used, to varying degrees, in different parts of the nervous system, and that both contribute to the ability of the nervous system to discriminate complex objects and produce graceful movements.

Information processing in sensory systems

Nearly 60 years ago, Claude Shannon developed a theory regarding the information that can be transmitted in a noisy communication channel⁴. This information theory has been popularized in terms of the 'bits and bytes' of information in computers, and has even been applied to the information content of the universe⁵. Application of the theory to neural systems seemed logical, as the survival of an organism depends on its ability to rapidly gather crucial information in a noisy or uncertain environment. One approach suggested that the presence (1) or absence (0) of a spike would represent one binary choice, which is commonly referred to as a binary digit (or 'bit') of information^{6,7}. For example, the binary numbers 00, 01, 10 and 11

^{*}Department of Physiology,
[‡]Centre for Neuroscience
and [§]Department of
Biomedical Engineering,
University of Alberta,
Edmonton,
Canada T6G 2S2
Correspondence to R.B.S.
e-mail: richard.
stein@ualberta.ca
doi:10.1038/nrn1668

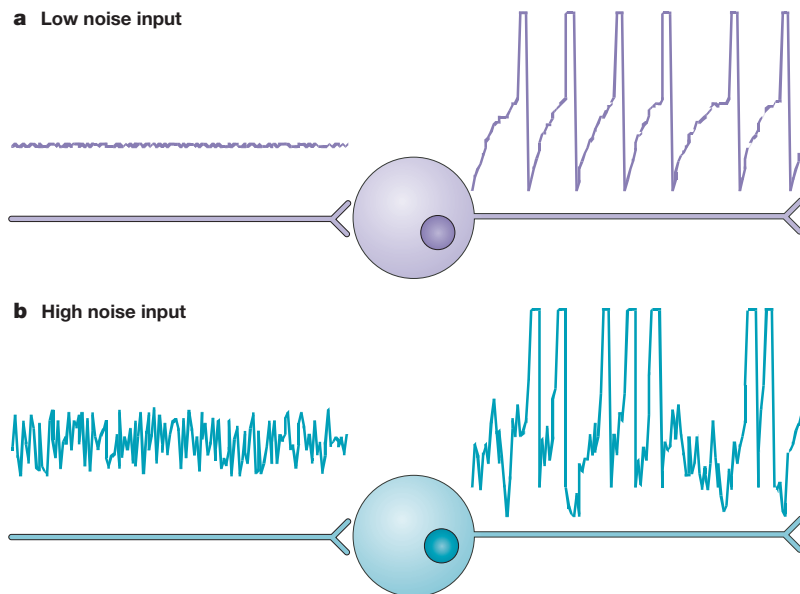


Figure 1 | Variability in neuronal firing. **a** | With a relatively steady depolarizing input current (low noise), spikes are generated at a regular rate. **b** | With higher noise, which could arise from a combination of excitatory and inhibitory postsynaptic potentials, the variability in firing rate is much higher, even though the mean firing rate might be similar. These data were simulated from a neural model with a leaky integrator and a fixed threshold⁸². The range in variability shown is typical of many neurons.

represent 2 bits (or $2^2 = 4$ choices). Therefore, if the REFRACTORY PERIOD of a neuron is less than 1 ms, it could transmit more than 1,000 (10^3) bits of information in 1 s in a sensory process. However, if different stimulus intensities are presented to a human observer (with 10^{12} neurons), and the observer assigns a numerical score to each stimulus presentation, approximately seven categories can be reliably discriminated⁸. When more choices are presented, observers make errors and the information calculated from information theory is still roughly equivalent to seven categories. So, although the system has the potential to transmit $10^3 \times 10^{12} = 10^{15}$ bits in 1 s, conscious perception can provide less than 3 bits (or $2^3 = 8$ categories) of information about stimulus intensities in a range of sensory systems.

The rate code. This large discrepancy was soon resolved experimentally and theoretically. As mentioned above, Lord Adrian proposed the rate code in which the intensity of a signal is represented as the number of spikes from a single neuron or a population of neurons over a period of time. The rate is obtained by dividing the number of spikes by the time period. In other words, signals in the sequence 100 (spike, no spike, no spike) over a period of 3 ms would be interpreted not as the binary representation of the number 4, but as a rate of 1 spike in 3 ms, which corresponds to 333 spikes in 1 s. This code would be indistinguishable from the sequences 010 and 001. In such a rate code, the total number of spikes over a period of time is the determining factor, rather than the order or timing of spikes. Variability in the inter-spike intervals and, therefore,

in the total number of spikes in 1 s, for example, would limit the amount of sensory information that neurons transmit.

If the input to a neuron has little noise (FIG. 1a), the neuron will depolarize to the threshold at a steady rate and the number or rate of spikes over a period of time will be reproducible. With a noisy input (FIG. 1b), the spikes will be generated with variable intervals and the rate of spikes over time will also vary. Values of information capacity of 2 or 3 bits have been calculated for single neurons in sensory systems, using rate coding to distinguish the levels of response to steady inputs over 1 s (REF. 9,10). In fact, Talbot and colleagues¹¹ directly calculated the information contained in the firing of sensory receptors of a monkey to a flutter vibration applied to the skin and found that it agreed with the information from human observers responding to the same sensory stimulus. This agreement between responses in monkey sensory receptors and human experiments was encouraging, but several questions remained.

The problems with the rate code. First, as most sensory stimuli activate many receptors, a human observer should be able to use the larger number of nerve impulses from the population of sensory neurons to distinguish more categories. In other words, if a whole population of neurons is active, and each neuron sends 2 or 3 bits to the brain in 1 s, why is human perception limited to less than 3 bits? Is the agreement between the calculated values of information capacity for single neurons and the human observer merely fortuitous? The answer to the latter question is probably yes. With a rate code, the number of bits of information increases according to the square root of the number of impulses, regardless of whether the impulses come from a single neuron or a population of neurons¹⁰. Therefore, 100 neurons would transmit 10 times the information of a single neuron. If the brain uses 100 neurons to perceive different categories, we would expect ~20–30 bits of information to be transmitted in 1 s, which predicts more than a million possible perceptual categories. The information contained in the responses from populations of cells must underlie the ability to recognize faces, as mentioned above. However, some information might be lost in synaptic relays or in memory storage in the brain. Conceivably, this loss might cancel out the extra information that is contained in the population of sensory neurons in the particular laboratory experiments quoted above and leave 2 or 3 bits (or 7 categories). These issues still need to be rigorously tested experimentally.

Second, if variability in inter-spike intervals creates noise that limits information flow in sensory systems, why have sensory neurons not evolved to reduce this variability? As discussed below, in relation to temporal codes, neurons do fire spikes at precise times in some sensory systems. In others, the receptors are operating close to physical limits that introduce variability. For example, in the visual system, photons arrive randomly in time and, under some conditions, a single photon

REFRACTORY PERIOD
The period of time after a spike when a neuron is unable or less able to fire another spike.

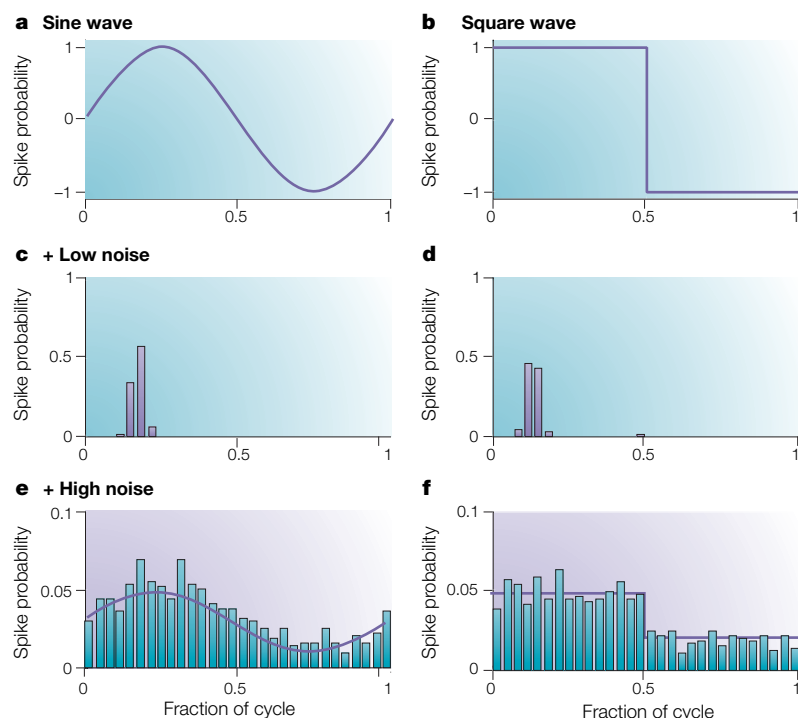


Figure 2 | Noise can be beneficial to the faithful transmission of high-frequency inputs. When a sine wave (a) or square wave (b) with a period of 30 ms is added as an input to the low-noise neural model shown in FIG. 1a, the input entrains the firing of the neural model so that it generates one spike per cycle at a relatively fixed phase (c and d) and the shape of the input is lost. Adding the same inputs to the high-noise model shown in FIG. 1b produces spikes at all phases and the probability follows the shape of the input (e and f).

can lead to perception¹². The variability in the timing of spikes might be related to the irregular arrival of photons. Similarly, Brownian motion affects hair cells in the auditory system^{13,14}. Therefore, the presence of variability in sensory systems might be an inevitable consequence of exquisitely responsive sensors. Furthermore, during synaptic transmission through sensory and central synapses, the effects of individual excitatory postsynaptic potentials (EPSPs) are relatively large compared with those of individual ions in the constant currents that were considered in FIG. 1a (REF. 15). If the generation of EPSPs is a random or STOCHASTIC PROCESS, variability will be introduced in the time that is taken to reach the threshold for spike generation^{16,17}. According to this view, the response variability of neurons in the CNS is a property of synaptic connections rather than the neurons themselves¹⁸. Below, we discuss evidence from the work of Mainen and Sejnowski, and others on the advantages of signal variability, which shows that the generation of EPSPs is not random and that their large size can be used to preserve timing information across synapses.

Finally, are these low rates of information transfer a consequence of studying steady signals? The answer to this question is probably also yes. The most important biological signals are changes in environmental parameters, such as light intensity. Sensory neurons respond to changing signals over a range of frequencies (BANDWIDTH) and can only signal information in their normal

working range. The bandwidth limits the maximum information capacity, but much more information can be transmitted with changing, rather than steady, signals. Experimental attempts to measure information capacity with broad-bandwidth random inputs have yielded approximate values of 1 bit per spike^{19,20}. If neurons fire tens or even hundreds of spikes per second, then tens or hundreds of bits per second are also possible, rather than only three with steady signals. Indeed, with rapidly varying signals, rates must be measured over small time intervals, so the distinction between rate and temporal codes breaks down. A more meaningful measure is the accuracy in the timing of spikes in individual neurons in response to the changing stimulus, which can be of the order of milliseconds in cortical neurons and tenths of a millisecond in some sensory neurons^{21,22}.

Advantages of signal variability. Recent studies indicate that variability might also offer distinct advantages. Noise could enhance sensitivity to weak signals, a phenomenon that is known as 'stochastic resonance'^{23–25}. As sensory signals are variable, Knill and Pouget²⁶ suggested that the brain might also code sensory information probabilistically and use the method of Bayesian inference. With this approach, the decision processes in the brain could deduce the best choice by combining previous experience with the probabilistic sensory signal²⁷.

An example of the potential advantages of signal variability is presented in FIG. 2. If a sine wave with a period of 30 ms is applied to a muscle or a cutaneous receptor (FIG. 2a), the neuron becomes ENTRAINED to the stimulus, with one spike for each sinusoidal cycle. This gives information about the period, but not the form of the stimulus. A square wave with the same period (FIG. 2b) produces a similar train of spikes. Variability in the individual neurons can prevent this entrainment^{28,29}. With little noise, using the neural model of FIG. 1a, the responses to the sine wave would occur near a specific point in the cycle and would be indistinguishable for a sine or square wave (FIG. 2c,d). With more noise, as typically occurs in many neurons, the inter-spike intervals are more variable and the cycle histograms assume distinct shapes. The average responses to the sine and the square waves now match the waveforms of the input signals (FIG. 2e,f).

The problem of entrainment becomes greater the higher the frequency of the applied signals. Therefore, it is most acute in the auditory system, as this receives tones with frequencies of several kilohertz, which is higher than in other sensory systems. However, a 400-Hz tone on a violin can be distinguished from a similar tone on an oboe. Although the fundamental frequency is the same, the HARMONICS that are produced by an oboe and a violin at 800 Hz and higher frequencies have different strengths. As the firing of cochlear neurons is close to that expected for a random or Poisson process³⁰, the resulting high degree of variability ensures that the different harmonic structures of the signals from the violin and oboe produce a different average

STOCHASTIC PROCESS

A random sequence of events; if the probability of occurrence of the events is the same with each small increment of time, it is referred to as a Poisson process.

BANDWIDTH

The range between the lowest and highest frequencies of oscillation that produce a response.

ENTRAINED

The state in which one signal is linked to the repetitive behaviour of another.

HARMONICS

Integral multiples of the fundamental frequency.

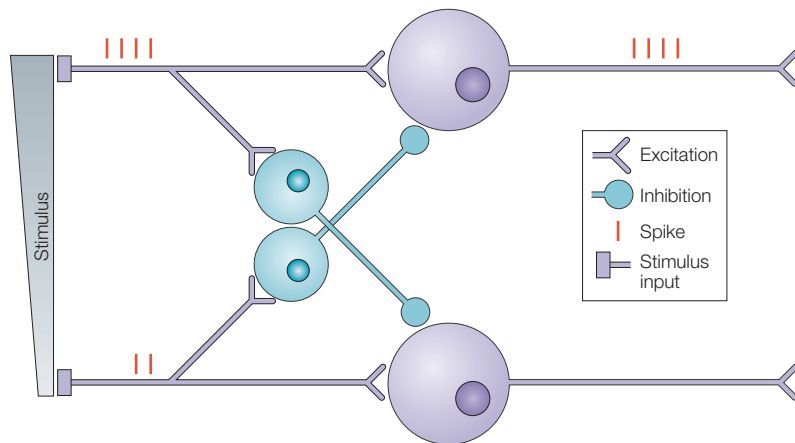


Figure 3 | Neuronal connections that are mutually inhibitory can accentuate the differences in time and intensity. Neuronal pathways (top) that are closer to a stimulus might fire earlier and produce more spikes than other pathways (bottom). Neuronal connections that are mutually inhibitory can enhance this difference by blocking transmission in neurons that receive later and less frequent excitation. Recent work has emphasized the importance of the latency to the first spike in rapid behavioural responses.

response. Therefore, the variability can increase the fidelity of transmission for high-frequency signals. In other words, although the variability is not necessarily part of the sensory signal, it might be an important part of the faithful processing of the signal. The ability of the neuron to transmit signals faithfully is only evident after analysing many cycles of the stimulus. However, transmission by a population of neurons, rather than a single neuron, would allow the signal to be evident in real time. There are also neurons in the population that are tuned to respond best to 800 Hz and higher frequencies; responses from these neurons might also be important in distinguishing the tones from a violin and an oboe. Therefore, recording from a population of neurons is crucial for understanding the functions of the sensory system.

Mainen and Sejnowski³¹ suggested another important advantage of noise. When they applied inputs with low noise, as in FIG. 1a, the timing of the spikes 'drifted' from one trial to the next. The rate, but not the exact timing, of the spikes was reproducible. The authors showed that the addition of EPSPs produced spikes reliably on a millisecond time scale in one or more neurons. As the EPSPs that were applied presynaptically reliably triggered spikes, the timing of presynaptic events could be faithfully transmitted to postsynaptic neurons. Mainen and Sejnowski used cortical neurons in their study, but this argument also applies to synaptic transmission in sensory pathways (for example, see the evidence presented by Kara and colleagues³²). In short, the apparent variability of neurons in spike generation represents signals from presynaptic EPSPs that provide important timing cues for their postsynaptic partners. This is often referred to as a 'temporal code'. Various authors have used this term differently, but the central idea is that the exact timing of nerve spikes conveys more information than the rate alone. As there is no absolute time

reference in the nervous system, the information is carried either in terms of the relative timing of spikes in a population of neurons or with respect to an ongoing **BRAIN OSCILLATION**. The accuracy of timing within a millisecond might be required to determine when and where a stimulus occurred. Several cues, such as intensity, timing and phase, are available to localize an auditory source in space, but recent commentaries have emphasized the timing cue, with titles such as 'Timing is everything'³³, 'Time is precious'³⁴ and 'Spike times make sense'³⁵.

The neural timing mechanisms for localization have been well studied in animals, including birds³⁶, mammals^{37,38} and electric fish³⁹. They involve **PHASE-LOCKED SIGNALS** from two different sources converging onto cells that serve as comparators. Mutually inhibitory connections in a neural circuit are important in sharpening the time and intensity differences. Neurons that are closer to a stimulus might receive more and earlier inputs than more distant neurons (FIG. 3). Mutual inhibition from the more strongly activated neurons could then block transmission in the more weakly activated ones. Recent studies have indicated that the precise timing of spikes is crucial in the function of the somatosensory⁴⁰ and visual systems^{41,42}. Johansson and Birznieks studied the neural mechanisms that are involved in avoiding slippage of an object that is grasped by the fingers of a human hand⁴⁰. They argued that the timing of the first spikes in a population of cutaneous neurons codes the direction of movement on the skin, which is essential for the natural manipulation of objects, and for a fast and accurate response to the onset of a slip. The rate coding of later spikes might signal the level of force that is required to hold the object. This study highlights an important point: temporal and rate codes are not mutually exclusive and can be used in the same sensory neurons. Other evidence indicates that the temporal information in sensory neurons is faithfully transmitted to the sensory cortex^{32,42,43,45}. In the rat barrel cortex, the timing of individual spikes carries 82–85% of the total information regarding stimulus localization and the timing of the first spike is crucial⁴⁵. In summary, although earlier studies emphasized the importance of the rate code and inferred that variability in spike timing reflected neural noise, recent studies, such as those quoted above, emphasize the importance of temporal codes for understanding information processing in sensory systems.

Information processing in motor systems

Similar to sensory systems, motor systems show both precision and variability in the timing of spike generation. Activity in the 'slip' detectors mentioned above leads to a reflex action with a short and precise latency in the spinal cord to resist the slip⁴⁶. At the level of the motor cortex, neurons have a preferred direction; that is, they respond best when the animal is planning to move in a particular direction^{47,48}. Others have argued that it is not a particular direction that is coded but a synergistic group of muscles⁴⁹,

BRAIN OSCILLATION
Rhythmic activity that can be recorded using electroencephalogram methods and that is usually divided into categories that are based on frequency; for example, the θ -rhythm is 4–8 Hz.

PHASE-LOCKED SIGNALS
When two (or more) periodic signals become linked at a particular part of the periodic cycle.

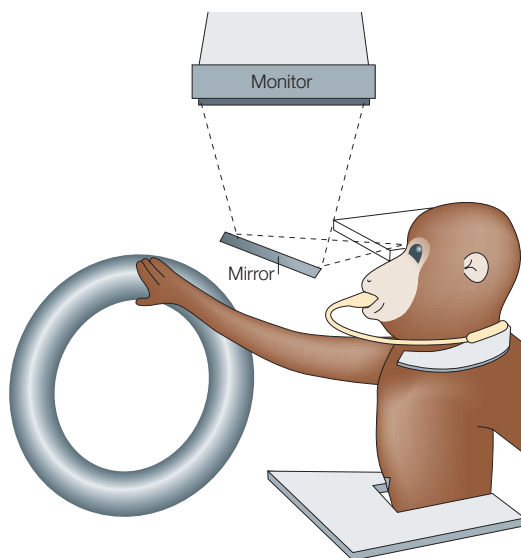


Figure 4 | Studying movements in virtual reality. In this experimental set up, the monkey is seated and views the three-dimensional image on the computer monitor that is projected onto the mirror in front of it. A marker on the hand causes a sphere on the monitor to provide an image of the otherwise obscured hand to the monkey. The template to trace (for example, an ellipse or a circle) is also projected for the monkey to see in the mirror. The monkey receives a juice reward through the feeding tube if it keeps the sphere within the template while tracing in virtual reality. Image modified with permission from REF. 83 © (2003) Elsevier Science.

or that movements and muscles are both coded, but by different cells⁵⁰. In any case, the response is generally variable and the neurons respond broadly with increased or decreased rates to movements in a wide range of directions. Therefore, the firing of a single neuron cannot accurately predict in which direction the animal will move. Schwartz and colleagues⁴⁸ trained monkeys to produce movements in virtual reality (FIG. 4). The activity of cells in the arm area of the motor cortex was recorded while the monkey made elliptical movements (FIG. 5a). The mean firing rate of a single neuron increased and decreased sinusoidally (FIG. 5b), but the variability in spike generation also increased as the mean rate increased (FIG. 5c). This study illustrates the importance of rate coding in the motor cortex. However, because of the variability, the number of neurons that would be required for a disabled subject to accurately control a movement using a neural prosthesis is estimated to range from tens to thousands of cells^{51,52}.

The optimal control theory. Harris and Wolpert⁵³ proposed an elegant OPTIMAL CONTROL THEORY in which movements are organized so as to minimize the effects of variability at the end point. Movements from one point to another can be made with different trajectories, speeds and accuracies. The best strategy for controlling a movement so that it reaches a desired end point with the least error depends on the variability that is inherent in the neural signals. If the variability

is known at different signal levels, the optimal control theory can be used to calculate the best activation patterns of relevant muscles to minimize error. In other words, variability in the motor system is considered as noise that decreases movement accuracy. However, if the nature of the noise is predictable, then the motor system might use optimal control algorithms to make the most accurate movements that are possible to a given target^{54,55}. This is a potential solution to the problem that was raised many years ago by John von Neumann⁵⁶, who was a founder of artificial intelligence: how can a reliable nervous system be made out of unreliable elements? Harris and Wolpert argued further that if the variability is predictable and obeys a particular rule, many properties of stereotypical motor behaviours will follow automatically. The required rule is that the standard deviation (SD) of the control signal varies proportionally to the mean control signal; in other words, the noise is 'signal dependent'. This rule would unify many experimental observations that were previously described in isolation without any common theoretical foundation, which is the hallmark of a good theory.

Does the Harris and Wolpert rule hold experimentally? One difficulty in answering this question is that they did not clearly specify what 'the control signal' meant in terms of the nervous system. *A priori*, it might be assumed that neurons in the motor cortex that are involved in planning the movement will also fire so as to minimize errors. If the firing rate of neurons in the motor cortex is the control signal, then the rule does not hold experimentally. As mentioned previously (FIG. 5), the mean rate of spikes in individual neurons increases and decreases sinusoidally as the monkey traces each cycle of the ellipse. The SD of the rate also changes and the logarithm of the SD in firing rate is plotted against the logarithm of the mean rate (FIG. 5c) for the neuron that is shown in FIG. 5b. The gradient of the slope is 0.48, which indicates that the SD of the firing rates increases approximately according to the square root of the mean rate (which produces a slope of 0.5 on this plot), rather than linearly (which would produce a slope of 1.0). FIGURE. 5d shows the results for 286 cells from 4 cortical hemispheres of 2 monkeys. Each cell was plotted as in FIG. 5c and the slope on this plot corresponds to the exponent of the relation between the SD of the firing rate and the mean rate. The Harris and Wolpert theory would predict a value of 1.0 if the relevant 'control signal' were the firing rate of neurons in the motor cortex. A range of values is seen in FIG. 5d, but the mean value of the slope is 0.50 ± 0.12 , which is significantly less than 1.0 ($P < 0.001$, student's *t*-test).

Perhaps the control signal is not the firing rate from single cortical neurons, but from populations of neurons. However, the same result with a value of 0.5 was also obtained with a population of cortical neurons (R.B.S., A.B. Schwartz and D.W. Moran, unpublished observations). Alternatively, the control signal might be derived from motor neurons in the spinal cord, rather than from neurons in the motor cortex.

OPTIMAL CONTROL THEORY
In engineering terms, a mathematical theory that allows for the regulation of a dynamic system using *a priori* knowledge or a model of the system to minimize particular variables, such as errors.

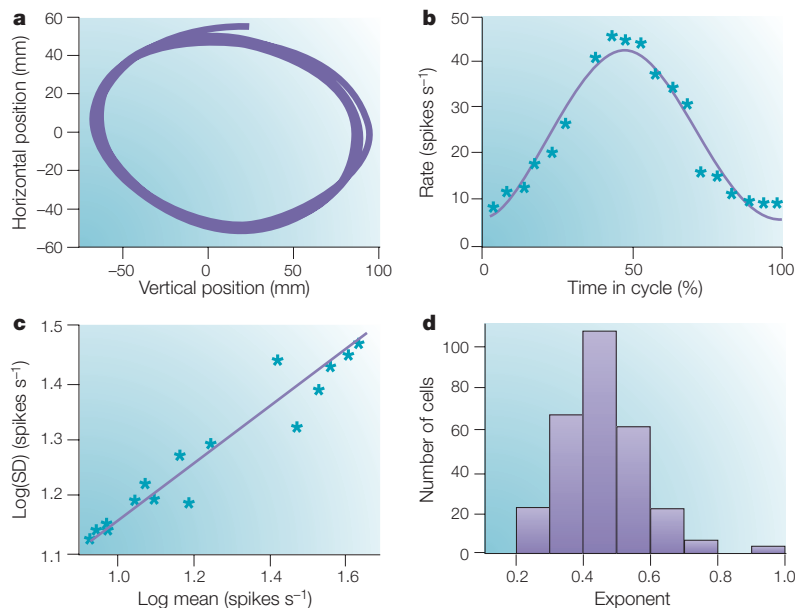


Figure 5 | Neural activity while drawing in virtual reality. **a** | The average trajectory in the frontal plane of three cycles of ellipses that were drawn in space by a monkey using the virtual reality display shown in FIG. 4. **b** | The mean firing rate of a neuron in the motor cortex is plotted against the percentage of the average time taken to draw one cycle of the ellipse (0.98 s). **c** | The mean and standard deviation of the firing rate of the neuron in each of 20 cycles plotted on log/log coordinates. The line was fitted using linear least mean-squares techniques. **d** | For each cell, the slope on the log/log plot of **c** was calculated, which corresponds to an exponent on a linear plot. This histogram shows the number of cells having various exponents, and was used to test a possible basis for the Harris and Wolpert optimal-control theory, as described in the text.

Harris and Wolpert quoted studies in which the SD of the inter-spike interval distributions for motor neurons was proportional to the mean interval. However, the control signal is more likely to be the rate of spikes, which is directly related to the force output, rather than the inter-spike intervals. BOX 1 shows that, even if the SD of the inter-spike intervals is proportional to the mean interval, as quoted in Harris and Wolpert, the SD of the number of spikes or the rate over a given time will vary as the square root of the mean. This is in agreement with the results from the cortex (FIG. 5), but not with the rule proposed by Harris and Wolpert. A final possibility is that the relevant control signal is not derived from the nervous system, but is the total force output of a muscle or a group of muscles. Several studies have now shown that the SD of the force is approximately proportional to the mean force level during voluntary ISOMETRIC CONTRACTIONS^{57–59}. How can the linear relationship between the SD of the force and the mean force arise from the square-root relationship in the nervous system? The answer to this question follows from the HENNEMAN size principle.

The Henneman size principle. Henneman proposed that the size of a motor neuron is essential to the structured relationship between its excitability threshold and the number, size and fatigability of the muscle fibres that it innervates⁶⁰. To oversimplify:

if changes in membrane voltage follow Ohm's law ($V = IR$), the threshold for generating a spike requires the same change in voltage (V) for all motor neurons, and the synaptic currents (I) from muscle receptors or descending voluntary inputs are the same in all motor neurons, then the current will depolarize the cell membrane of a small neuron, which has a high input resistance (R), faster than that of a large neuron. As a result, the threshold will be reached sooner and more often. Therefore, motor neurons that produce small forces are the first to undergo RECRUITMENT and are followed by those that produce larger forces⁶¹. Although these assumptions are not entirely correct, the errors that are introduced do not change the basic recruitment order of motor neurons from those that produce small forces to those that produce large forces^{62–64}. A MOTOR UNIT (FIG. 6c) that is active in a small reflex or voluntary contraction is small and non-fatiguing and the force fluctuations are also relatively small. As the strength of contraction increases, progressively larger and more fatigable motor units are recruited. This 'size principle' has been studied in several muscles in animals and humans^{62,63}, and a model is available for the first dorsal interosseous muscle of the human hand⁶⁴.

The relationship between the SD of the force and the mean force level can be calculated under the assumption that the motor units are either recruited in a random order (FIG. 6a) or are recruited according to size in an orderly way (FIG. 6b). In recruitment according to size (FIG. 6c), the smallest motor unit, which has the lowest threshold, is the most likely to respond to the input. Given that the thresholds gradually increase as the size of the motor units increase, there is a reduced chance of response to the input and fewer action potentials are produced in the larger motor units. Under the assumption of random recruitment, the SD increases according to the square root of the mean force level. However, if the motor units are recruited according to size (FIG. 6d), the SD is smaller, particularly at low force levels, and increases linearly with the mean force level. In other words, according to the Henneman size principle, small motor units are initially recruited to fire at low force levels that together produce relatively small fluctuations in force. Progressively larger motor units are recruited at higher force levels, which inevitably produce larger fluctuations. This results in a steeper and more linear relationship between the SD and the mean force level, which is consistent with the rule required by the Harris and Wolpert theory.

Box 1 | Variability in intervals and rate

If μ is the mean inter-spike interval and σ is its standard deviation (SD), then the mean number of spikes, M , in time T will be $M = T/\mu$ and the variance in number, V , will be approximately $V = \sigma^2 T/\mu^3$ (REF. 81). If the coefficient of variation, σ/μ , is a constant, C , then the SD in number will be $SD = C\sqrt{M}$.

ISOMETRIC CONTRACTION
A contraction in which a muscle exerts force but does not change in length.

HENNEMAN
Elwood Henneman (1915–1995) was a neuroscientist at Harvard Medical School, who proposed a theory about the functional significance of cell size in spinal motor neurons.

RECRUITMENT
With respect to motor units, when a previously inactive unit is excited beyond its threshold and begins producing action potentials.

MOTOR UNIT
A motor neuron and all the muscle fibres that it innervates.

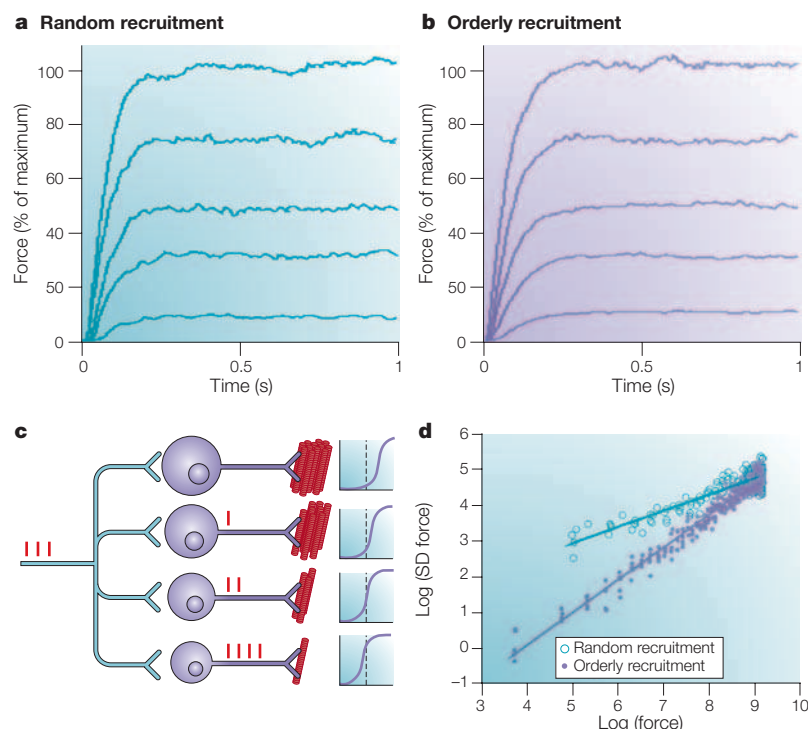


Figure 6 | Simulations of force variability in a model of a motor pool. **a** | If the threshold for recruitment of a pool of 120 motor units is random with respect to twitch force, considerable variability occurs even during sustained contractions at low levels of mean force. As the level of mean force increases, the variability around the mean increases further. **b** | Normally, motor units in a pool are recruited in order of their twitch force, from small to large. With orderly recruitment, the variability is less at the same level of mean force compared with random recruitment. **c** | A schematic diagram of the orderly recruitment of motor units. A motor unit consists of a motor neuron and all of the muscle fibres it innervates. A common input drives all four motor neurons, but the combination of active and passive membrane properties results in thresholds (represented by the sigmoidal curves on the right) that vary across the pool of motor neurons^{56,76}. The smallest motor unit has the lowest threshold current (sigmoid shifted to the left) and is therefore most likely to respond to the input. The threshold functions are gradually shifted to the right for the larger motor units, which results in a decreased probability of response to the input and fewer action potentials being sent to their constituent muscle fibres. Red vertical bars represent spikes. **d** | In both types of recruitment, variability increases as a function of mean force output, but at a different rate. Regression analysis gave a slope of 0.45 on logarithmic scales during random recruitment (open circles; $r^2 = 0.77$, $F = 820.74$, $P < 0.001$) and a slope of 0.90 during orderly recruitment (filled circles; $r^2 = 0.97$, $F = 6.5429$, $P < 0.001$). Details of the model are given in REF. 57.

In contrast to the sensory system, in which variability can be beneficial, inconsistency in the motor system is an unavoidable concomitant of motor output. However, because the variability is both predictable and linear, the motor system might have evolved optimal methods for generating accurate movements despite the ‘noise’. The trajectories of movements can then be organized to minimize errors in the end point. It should be possible to implement these same methods in a neural prosthesis to optimize the control of movements. How the nervous system calculates the optimal paths remains to be determined, but natural movements seem to follow the predicted trajectories, even when avoiding obstacles⁶⁵. In summary, recent studies on the motor system indicate that the stereotypical structure of movements might result, at least in part, from the attempts to minimize

the effects of inherent variability or noise in the force output of muscles.

Central processing of information

Do other regions of the CNS use similar principles to deal with variability or is the cognitive processing in these regions fundamentally different from the sensory-motor events discussed above? Konig and colleagues⁶⁶ suggested that the way in which a neuron processes information depends on the time period over which it integrates the input signals⁶⁶. If the period is long compared with an inter-spike interval, then the firing rate will signal the integrated inputs that the neuron receives. However, if the period is short, the neuron will only fire when inputs are synchronized and will function as a COINCIDENCE DETECTOR. The authors argued that cortical neurons predominantly function as coincidence detectors⁶⁶. Stevens and Zador⁶⁷ injected current, which was intended to mimic noisy synaptic current, into neocortical neurons using a microelectrode in brain slices. They found that the irregular firing typical of cortical neurons occurred only in response to large synchronous inputs. Steinmetz and colleagues⁶⁸ described an interesting example in the monkey somatosensory cortex. Neurons became increasingly synchronized when an animal was paying attention, particularly to difficult sensory discriminations. Increased synchronization of motor units was also observed during tasks that demanded attention. The synchronous activation of separate pathways has long been associated with the potentiation of synapses, which might be important for learning and memory^{69–71}; however, this topic is outside the scope of the present review.

Integration and synchronization are not always in opposition, as indicated by Konig and colleagues⁶⁶ and discussed above. The firing rate of the ‘place neurons’ in the hippocampus, increases when an animal is near its preferred location in space. However, the spikes are synchronized with respect to the θ -rhythm, which is prominent in the hippocampus, at a phase that changes monotonically as the animal approaches and then passes its preferred location^{72–74}. Although this PHASE INFORMATION is available to the researcher, who can record both spike and wave activity, it is still unclear whether cells downstream to the place neurons extract this phase information. Oscillations similar to the θ -rhythm are widespread in the nervous system, but their meaning for specific behaviours is still debated. Mehta and colleagues⁷⁵ suggested that oscillations might convert a rate code into a temporal code. They found that, in the rat hippocampus, the temporal code became more robust with experience and might be involved in learning temporal sequences. Tsodyks and colleagues⁷⁶ found that the spontaneous activity of individual neurons was strongly associated with the spatial patterns in the visual cortex of cats, so oscillations might serve to bind neurons into ensembles that function together.

COINCIDENCE DETECTOR

A sensing device that receives inputs from many sources and preferentially responds when these inputs arrive synchronously.

PHASE INFORMATION

The fraction of a complete cycle as measured from a specific reference point.

Concluding remarks

In summary, the variability in the inter-spike intervals of individual neurons is more than 'neural noise', although its full importance in cognitive processing remains to be determined. In both sensory and cortical systems, correlated activity can preserve the timing of events across synaptic relays between cells, and neural synchrony might be important in attention, learning and memory. In motor systems, rate coding might be more important than precise temporal coding. Variability needs to be reduced so that smooth accurate movements can be produced and

the optimal control theory indicates how this can be achieved. Since the publication of the Shannon theory of information⁴, much has been learnt about information processing in the nervous system, yet it remains an exciting topic for active research. New techniques should lead to considerable progress in the near future. In particular, the technique for simultaneously recording neuronal activity from many neurons will be crucial^{51,77–80} for understanding information processing in sensory systems as well as designing neural prostheses that are controlled by cortical signals.

1. Adrian, E. & Zotterman, Y. The impulses produced by sensory nerve endings. Part 3. Impulses set up by touch and pressure. *J. Physiol. (Lond.)* **61**, 465–483 (1926).
This paper provided the first evidence for rate coding in sensory neurons, which led to Lord Adrian receiving the Nobel Prize in Physiology or Medicine (1932).
2. Adrian, E. & Bronk, D. The discharge of impulses in motor nerve fibres. Part II. The frequency of discharge in reflex and voluntary contractions. *J. Physiol. (Lond.)* **67**, 119–151 (1929).
3. Khamsi, R. Paralyzed man sends e-mail by thought. *News@Nature* 13 October 2004 (doi:10.1038/news041011-9).
4. Shannon, C. A mathematical theory of communication. *Bell. System. Tech. J.* **27**, 379–423 (1948).
This classic monograph presented a complete theory of information transmitted through a noisy communication channel. It has been argued that this paper marked the beginning of the information age.
5. Lloyd, S. & Ng, Y. J. Black hole computers. *Sci. Am.* **291**, 52–61 (2004).
6. Rappoport, A. & Horvath, W. J. The theoretical channel capacity of a single neuron as determined by various coding systems. *Inf. Control* **3**, 335–350 (1960).
7. MacKay, D. & McCulloch, W. The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* **14**, 127–135 (1952).
8. Miller, G. The magical number seven plus or minus two. Some limits on our capacity for processing information. *Psych. Rev.* **63**, 81–97 (1956).
9. Werner, G. & Mountcastle, V. B. Neural activity in mechanoreceptive cutaneous afferents: stimulus–response relations, Weber functions, and information transmission. *J. Neurophysiol.* **28**, 359–397 (1965).
10. Stein, R. B. The information capacity of nerve cells using a frequency code. *Biophys. J.* **7**, 797–826 (1967).
11. Talbot, W. H., Darian-Smith, I., Kornhuber, H. & Mountcastle, V. B. The sense of flutter vibration: comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *J. Neurophysiol.* **31**, 301–334 (1968).
12. Pirenne, M. H. Some aspects of the sensitivity of the eye. *Ann. NY Acad. Sci.* **74**, 377–384 (1959).
13. Harris, G. G. Brownian motion in the cochlear partition. *J. Acoust. Soc. Am.* **44**, 176–186 (1968).
14. Corey, D. P. & Hudspeth, A. J. Kinetics of the receptor current in bullfrog saccular hair cells. *J. Neurosci.* **3**, 962–976 (1983).
15. Bekkers, J. M. & Stevens, C. F. Excitatory and inhibitory autaptic currents in isolated hippocampal neurons maintained in cell culture. *Proc. Natl Acad. Sci. USA* **88**, 7834–7838 (1991).
16. Stein, R. B. A theoretical analysis of neuronal variability. *Biophys. J.* **5**, 173–194 (1965).
17. Calvin, W. H. & Stevens, C. F. Synaptic noise and other sources of randomness in motoneuron interspike intervals. *J. Neurophysiol.* **31**, 574–587 (1968).
18. Movshon, J. A. Reliability of neuronal responses. *Neuron* **27**, 412–414 (2000).
19. de Ruyter van Steveninck, R. R. & Bialek, W. Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. *Proc. R. Soc. Lond. B* **234**, 379–414 (1988).
20. Borst, A. & Theunissen, F. E. Information theory and neural coding. *Nature Neurosci.* **2**, 947–957 (1999).
21. Arabzadeh, E., Zorzin, E. & Diamond, M. E. Neuronal encoding of texture in the whisker sensory pathway. *PLoS Biol.* **3**, 155–165 (2005).
22. Petersen, R. S., Panzeri, S. & Diamond, M. E. Population coding in somatosensory cortex. *Curr. Opin. Neurobiol.* **12**, 441–447 (2002).
23. Wiesenfeld, K. & Moss, F. Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDS. *Nature* **373**, 33–36 (1995).
24. Jaramillo, F. & Wiesenfeld, K. Mechano-electrical transduction assisted by Brownian motion: a role for noise in the auditory system. *Nature Neurosci.* **1**, 384–388 (1998).
25. Cordo, P. et al. Noise in human muscle spindles. *Nature* **383**, 769–770 (1996).
26. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).
27. Kording, K. P., Ku, S. P. & Wolpert, D. M. Bayesian integration in force estimation. *J. Neurophysiol.* **92**, 3161–3165 (2004).
28. Stein, R. B. In *The Neurosciences, Second Study Program* (ed. Schmitt, F. O.) 597–604 (Rockefeller Univ. Press, New York, 1970).
29. Schellart, N. A. & Spekreijse, H. Dynamic characteristics of retinal ganglion cell responses in goldfish. *J. Gen. Physiol.* **59**, 1–21 (1972).
This paper provided experimental evidence for a positive role of noise in sensory systems, which was largely overlooked until the idea was re-examined recently in references 24 and 25.
30. Kiang, N., Watanabe, T., Thomas, E. & Clark, L. *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, Massachusetts, 1965).
31. Mainen, Z. F. & Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* **268**, 1503–1506 (1995).
32. Kara, P., Reinagel, P. & Reid, R. C. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* **27**, 635–646 (2000).
33. Carr, C. E. Timing is everything: organization of timing circuits in auditory and electrical sensory systems. *J. Comp. Neurol.* **472**, 131–133 (2004).
34. Abeles, M. Time is precious. *Science* **304**, 523–524 (2004).
35. Vanrullen, R., Guyonneau, R. & Thorpe, S. J. Spike times make sense. *Trends Neurosci.* **28**, 1–4 (2005).
36. Konishi, M., Takahashi, T. T., Wagner, H., Sullivan, W. E. & Carr, C. E. In *Auditory Function: Neurological Bases of Hearing* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 721–745 (Wiley, New York, 1988).
37. Goldberg, J. M. & Brown, P. B. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* **32**, 613–636 (1969).
38. Forsythe, I. D. Direct patch recording from identified presynaptic terminals mediating glutamatergic EPSCs in the rat CNS, *in vitro*. *J. Physiol. (Lond.)* **479**, 381–387 (1994).
39. Matsushita, A. & Kawasaki, M. Unitary giant synapses embracing a single neuron at the convergent site of time-coding pathways of an electric fish, *Gymnarchus niloticus*. *J. Comp. Neurol.* **472**, 140–155 (2004).
40. Johansson, R. S. & Birznieks, I. First spikes in ensembles of human tactile afferents code complex spatial fingertip events. *Nature Neurosci.* **7**, 170–177 (2004).
This paper provides compelling evidence that the latency of the first spikes in neurons of the human somatosensory system provides important positional information. Reference 45 indicates that such information is transmitted to the cortex in rats.
41. Reich, D. S., Mechler, F. & Victor, J. D. Temporal coding of contrast in primary visual cortex: when, what, and why. *J. Neurophysiol.* **85**, 1039–1050 (2001).
42. Thorpe, S., Delorme, A. & Van Rullen, R. Spike-based strategies for rapid processing. *Neural Netw.* **14**, 715–725 (2001).
43. deCharms, R. C., Blake, D. T. & Merzenich, M. M. Optimizing sound features for cortical neurons. *Science* **280**, 1439–1443 (1998).
44. Lestienne, R. Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Prog. Neurobiol.* **65**, 545–591 (2001).
45. Petersen, R. S., Panzeri, S. & Diamond, M. E. Population coding of stimulus location in rat somatosensory cortex. *Neuron* **32**, 503–514 (2001).
46. Johansson, R. S. & Westling, G. Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Exp. Brain Res.* **56**, 550–564 (1984).
47. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
48. Schwartz, A. B. Direct cortical representation of drawing. *Science* **265**, 540–542 (1994).
49. Scott, S. H., Gribble, P. L., Graham, K. M. & Cabel, D. W. Dissociation between hand motion and population vectors from neural activity in motor cortex. *Nature* **413**, 161–165 (2001).
50. Kakei, S., Hoffman, D. S. & Strick, P. L. Muscle and movement representations in the primary motor cortex. *Science* **285**, 2136–2139 (1999).
51. Taylor, D. M., Helms-Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
This paper showed that, with a learning algorithm, relatively few cortical neurons in a monkey could generate improved real-time control of movements in three dimensions. It provided a proof-of-principle for the implantation of electrode arrays in the motor cortex of severely disabled humans, as mentioned in the text.
52. Nicolelis, M. A. Brain-machine interfaces to restore motor function and probe neural circuits. *Nature Rev. Neurosci.* **4**, 417–422 (2003).
53. Harris, C. M. & Wolpert, D. M. Signal-dependent noise determines motor planning. *Nature* **394**, 780–784 (1998).
This paper provides a theoretical framework based on optimal control that explains a wide range of experimental observations on motor behaviour. Our review examines the extent to which their assumptions are experimentally justified.
54. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5**, 1226–1235 (2002).
55. Scott, S. H. Optimal feedback control and the neural basis of volitional motor control. *Nature Rev. Neurosci.* **5**, 532–546 (2004).
56. von Neumann, J. In *Automata Studies* (eds Shannon, C. & McCarthy, J.) 43–98 (Princeton Univ. Press, Princeton, New Jersey, 1956).
57. Jones, K. E., Hamilton, A. F. & Wolpert, D. M. Sources of signal-dependent noise during isometric force production. *J. Neurophysiol.* **88**, 1533–1544 (2002).
58. Hamilton, A. F., Jones, K. E. & Wolpert, D. M. The scaling of motor noise with muscle strength and motor unit number in humans. *Exp. Brain Res.* **157**, 417–430 (2004).
59. Taylor, A. M., Christou, E. A. & Enoka, R. M. Multiple features of motor-unit activity influence force fluctuations during isometric contractions. *J. Neurophysiol.* **90**, 1350–1361 (2003).

60. Henneman, E. Relation between size of neurons and their susceptibility to discharge. *Science* **126**, 1345–1347 (1957).
This paper established the orderly recruitment of motor neurons in reflex contractions, which was later extended to human voluntary contractions in reference 62.
61. Gustafsson, B. & Pinter, M. J. On factors determining orderly recruitment of motor units: a role for intrinsic membrane properties. *Trends Neurosci.* **8**, 431–433 (1985).
62. Milner-Brown, H. S., Stein, R. B. & Yemm, R. The orderly recruitment of human motor units during voluntary isometric contractions. *J. Physiol. (Lond.)* **230**, 359–370 (1973).
63. Desmedt, J. E. & Godaux, E. Ballistic contractions in man: characteristic recruitment pattern of single motor units of the tibialis anterior muscle. *J. Physiol. (Lond.)* **264**, 673–693 (1977).
64. Fuglevand, A. J., Winter, D. A. & Patla, A. E. Models of recruitment and rate coding organization in motor-unit pools. *J. Neurophysiol.* **70**, 2470–2488 (1993).
65. Hamilton, A. F. & Wolpert, D. M. Controlling the statistics of action: obstacle avoidance. *J. Neurophysiol.* **87**, 2434–2440 (2002).
66. Konig, P., Engel, A. K. & Singer, W. Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* **19**, 130–137 (1996).
67. Stevens, C. F. & Zador, A. M. Input synchrony and the irregular firing of cortical neurons. *Nature Neurosci.* **1**, 210–217 (1998).
68. Steinmetz, P. N. *et al.* Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404**, 187–190 (2000).
69. McNaughton, B. L., Douglas, R. M. & Goddard, G. V. Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. *Brain Res.* **157**, 277–293 (1978).
70. Barrionuevo, G. & Brown, T. H. Associative long-term potentiation in hippocampal slices. *Proc. Natl Acad. Sci. USA* **80**, 7347–7351 (1983).
71. Levy, W. B. & Steward, O. Synapses as associative memory elements in the hippocampal formation. *Brain Res.* **175**, 233–245 (1979).
72. Skaggs, W. E., McNaughton, B. L., Wilson, M. A. & Barnes, C. A. θ -phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* **6**, 149–172 (1996).
73. O'Keefe, J. & Recce, M. L. Phase relationship between hippocampal place units and the EEG θ -rhythm. *Hippocampus* **3**, 317–330 (1993).
This paper described how the relationship between spike and slow rhythms could be used to code information about the place of an animal with respect to a preferred location.
74. Harris, K. D. *et al.* Spike train dynamics predicts θ -related phase precession in hippocampal pyramidal cells. *Nature* **417**, 738–741 (2002).
75. Mehta, M. R., Lee, A. K. & Wilson, M. A. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature* **417**, 741–746 (2002).
76. Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943–1946 (1999).
77. Buzsaki, G. Large-scale recording of neuronal ensembles. *Nature Neurosci.* **7**, 446–451 (2004).
78. Chapin, J. K., Moxon, K. A., Markowitz, R. S. & Nicolelis, M. A. L. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.* **2**, 664–670 (1999).
79. Hatsopoulos, N. G., Ojakangas, C. L., Paninski, L. & Donoghue, J. P. Information about movement direction obtained from synchronous activity of motor cortical neurons. *Proc. Natl Acad. Sci. USA* **95**, 15706–15711 (1998).
80. Stein, R. B. *et al.* Coding of position by simultaneously recorded sensory neurones in the cat dorsal root ganglion. *J. Physiol. (Lond.)* **560**, 883–896 (2004).
81. Cox, D. R. & Miller, H. D. *The Theory of Stochastic Processes* (Methuen & Co., London, 1965).
82. French, A. S. & Stein, R. B. A flexible neural analog using integrated circuits. *IEEE Trans. Biomed. Eng.* **17**, 248–253 (1970).
83. Reina, G. A. & Schwartz, A. B. Eye-hand coupling during closed-loop drawing: evidence of shared motor planning? *Hum. Mov. Sci.* **22**, 137–152 (2003).

Acknowledgements

We thank A. Schwartz and C. Dickson, as well as a panel of interested students, L. Major, P. Malik and Y. Mao for helpful suggestions on the manuscript.

Competing interests statement

The authors declare no competing financial interests.

Online links

FURTHER INFORMATION

Jones' homepage: <http://www.ualberta.ca/~kejones>

Stein's homepage: <http://www.physiology.ualberta.ca/Stein.htm>

Access to this interactive links box is free online.

Brain–machine interfaces: past, present and future

Mikhail A. Lebedev¹ and Miguel A.L. Nicolelis²

¹ Department of Neurobiology and Center for Neuroengineering, Duke University, Durham, NC 27710, USA

² Department of Biomedical Engineering and Department of Psychological and Brain Sciences, Duke University, Durham, NC 27710, USA

Since the original demonstration that electrical activity generated by ensembles of cortical neurons can be employed directly to control a robotic manipulator, research on brain–machine interfaces (BMIs) has experienced an impressive growth. Today BMIs designed for both experimental and clinical studies can translate raw neuronal signals into motor commands that reproduce arm reaching and hand grasping movements in artificial actuators. Clearly, these developments hold promise for the restoration of limb mobility in paralyzed subjects. However, as we review here, before this goal can be reached several bottlenecks have to be passed. These include designing a fully implantable biocompatible recording device, further developing real-time computational algorithms, introducing a method for providing the brain with sensory feedback from the actuators, and designing and building artificial prostheses that can be controlled directly by brain-derived signals. By reaching these milestones, future BMIs will be able to drive and control revolutionary prostheses that feel and act like the human arm.

Introduction

Less than a decade ago, hardly anyone could have predicted that attempts to build direct functional interfaces between brains and artificial devices, such as computers and robotic limbs, would have succeeded so readily, and in the process would have led to the establishment of a new area at the frontier of systems neuroscience. Born as a highly multidisciplinary field, basic research on brain–machine interfaces (BMIs) has moved at a stunning pace since the first experimental demonstration in 1999 that ensembles of cortical neurons could directly control a robotic manipulator [1]. Since then, a continuous stream of research papers has kindled an enormous interest in BMIs among the scientific community and the lay public. This interest stems from the considerable potential of this technology for restoration of motor behaviors in severely handicapped patients.

Indeed, BMIs have been primarily conceived as a potential new therapy to restore motor control in severely disabled patients, particularly those suffering from devastating conditions such as amyotrophic lateral

sclerosis (ALS), spinal cord injury, stroke and cerebral palsy. As this technology advances and the risks of invasive brain recordings decrease, BMIs might also hold promise for amputees. In addition to the systems controlling upper-limb prostheses, BMIs dedicated to the restoration of locomotion and speech are likely to emerge.

However, such stellar progress also breeds unrealistic expectations that such a future is just around the corner. Thus, the understandable eagerness in attaining the lofty goal of helping severely disabled patients has to be carefully calibrated by an objective analysis of the current state and future directions of the field. Such analysis indicates that, despite the optimism raised by a barrage of new accomplishments, there are still many issues that preclude straightforward translation of experimental BMIs into clinical applications. Indeed, most of the invasive BMIs have been tested only in experimental animals. Thus, despite recent enthusiasm to move emergent, and in some cases not thoroughly tested, BMI-related technology into clinical trials, much experimentation remains to be done before BMIs can become a safe and efficient rehabilitation tool.

Here, we highlight some of the fundamental obstacles faced by BMI research and propose a series of milestones that can transform recent experimental advances into viable clinical applications in the next 10–20 years. The roadmap detailed here takes into account the recent history of the field, the factors that influenced its growth, and a critical analysis of the published work.

Non-invasive BMIs

Figure 1 depicts a classification of the BMIs (or brain–computer interfaces, BCIs) developed during the past decade. The first feature that distinguishes BMIs is whether they utilize invasive (i.e. intra-cranial) or non-invasive methods of electrophysiological recordings. Non-invasive systems primarily exploit electroencephalograms (EEGs) to control computer cursors or other devices. This approach has proved useful for helping paralyzed or ‘locked in’ patients develop ways of communication with the external world [2–11]. However, despite having the great advantage of not exposing the patient to the risks of brain surgery, EEG-based techniques provide communication channels of limited capacity. Their typical transfer rate is currently 5–25 bits s^{−1} [2,11]. Although such a transfer rate might not be sufficient to control the movements of an arm or leg prosthesis that has multiple degrees of freedom, past and recent research in this field seems to

Corresponding author: Nicolelis, M.A.L. (nicoleli@neuro.duke.edu)
Available online 21 July 2006.

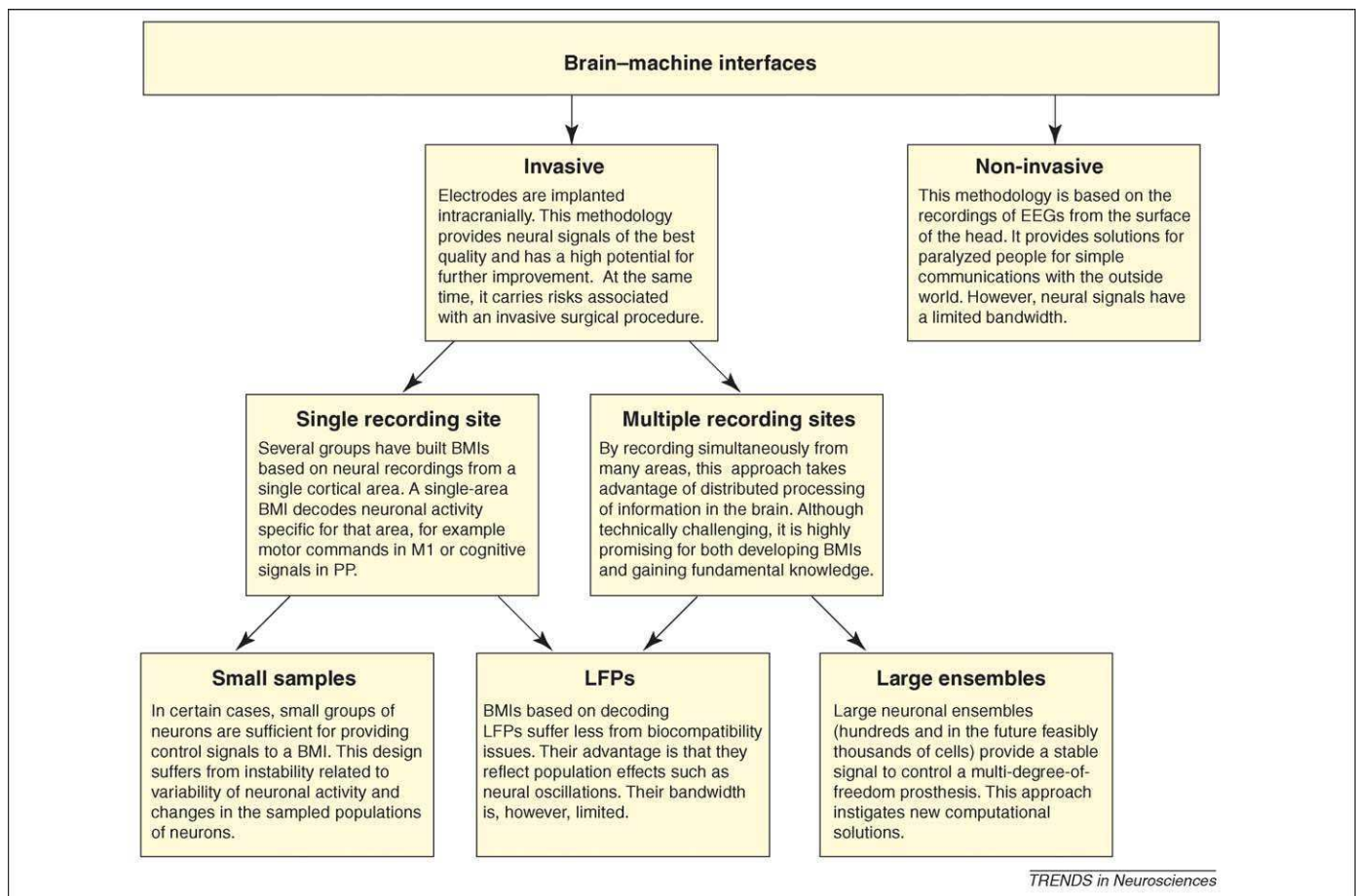


Figure 1. Classification of brain-machine interfaces. Abbreviations: BMI, brain machine interface; EEG, electroencephalogram; LFP, local field potential; M1, primary motor cortex; PP, posterior parietal cortex.

indicate that EEG-based BMIs are likely to continue to offer some practical solutions (e.g. cursor control, communication, computer operation and wheelchair control) for patients in the future.

Original attempts to provide subjects with feedback signals derived from their own brain activity were made in the 1960s and 1970s. Primarily, these attempts were aimed at enabling human subjects to gain voluntary control over brain rhythms. Nowlis and Kamiya claimed that, after training with an EEG biofeedback, human subjects acquired an ability to detect their own alpha rhythms [12], and this claim was later scrutinized by Plotkin [13]. Sterman and colleagues came to similar conclusions by utilizing the sensorimotor mu rhythm in cats [14] and humans [15]. In addition, Black operantly conditioned dogs to control their hippocampal theta rhythm [16]. Clearly, the results of these pioneering experiments helped pave the way for the introduction, a few years later, of EEG-based BCIs.

Generally, EEG-based BCIs try to decipher the subject's voluntary intentions and decisions through measurements of the combined electrical activity of massive neuronal populations. As such, both the spatial and temporal resolution of EEGs become limited owing to the overlapping electrical activity generated by different cortical areas. Furthermore, during the passive conductance of these signals through brain tissue, bone and skin, resolution is also lost owing to the low-pass filtering of the EEG signals.

EEGs are also susceptible to electromyographic (EMG), electrooculographic (EOG) and mechanical artifacts.

Despite these well-known shortcomings, EEG techniques can detect modulations of brain activity that correlate with visual stimuli, gaze angle, voluntary intentions and cognitive states. These properties have led to development of several classes of EEG-based systems, which differ according to the cortical areas recorded, the features of EEG signals extracted, and the sensory modality providing feedback to subjects. One class of BCIs makes use of visual evoked potentials (VEPs). These BCIs detect the VEPs that occur when subjects look at particular items on a computer screen [17,18] or attend to them [19]. BCIs based on the P300 evoked potential uncover the subjects' choices by distinguishing parietal cortex responses to the preferred versus non-preferred stimuli [20–22]. Several BCI designs continuously drive computer cursors. Both slow cortical potentials, recorded over several cortical areas [23], and faster mu (8–12 Hz) and beta (18–26 Hz) rhythms, recorded over sensorimotor cortex [24–26], have been exploited in such BCIs. For example, one such system relies on event-related synchronization and desynchronization of the EEGs associated with motor imagery [25,27].

Training to operate EEG-based BCIs can take many days [2]. Visual feedback is the essential part of such training. Some BCI designs rely on the subjects' ability to develop control of their own brain activity using biofeedback, whereas others utilize classifier algorithms that

recognize EEG patterns related to particular voluntary intentions. Recently, adaptive algorithms that constantly update the classifier parameters during training have been implemented [26].

Several strategies have also been proposed to provide feedback to users of EEG-based BCIs. For instance, virtual-reality systems can provide a realistic feedback that can be efficient for BCI training [28]. In a recent demonstration of this approach, subjects navigated through a virtual environment by imagining themselves walking [29].

In an effort to improve the resolution of brain potentials monitored by the BCIs, more invasive recording methods, such as electrocorticograms (ECoGs) recorded by subdural electrodes, have been introduced. ECoGs sample neuronal activity from smaller cortical areas than conventional EEGs. In addition, they contain higher-frequency gamma rhythms (>30 Hz). Consequently, ECoG-based BCIs are expected to have better accuracy and shorter training times than BCIs based on EEGs [30].

EEG-based BCIs have been implemented as solutions for patients suffering from various degrees of body paralysis. These BCIs (in the case of patients with advanced ALS) enable control of computer cursors, which the patients use to communicate with the external world or to indicate their intentions. The first successful and most well received application of such an approach was based on the utilization of slow cortical potentials to control a computer-aided spelling system [3,31].

BCIs based on mu and beta rhythms have also been tested in severely paralyzed people [32]. One study reported that a tetraplegic patient, aided by a BCI that detected beta waves in his sensorimotor cortex and activated a functional electrical stimulation device, learned to grasp objects using his paralyzed hand [33]. A motor imagery-based system [32], coupled to an implanted neuroprosthesis system [34] (Freehand[®]) has been used to help a partially paralyzed patient. In addition, tetraplegic patients were able to gain some degree of control of the P300-based BCI [21]. Off-line analyses showed that P300 potentials can be used to obtain information about stimulus selections made by patients with ALS [22].

In addition to using EEGs, imaging techniques such as functional magnetic resonance imaging (fMRI), have been explored as a new source of brain-derived signals to drive BCIs [35]. Although fMRI-based BCIs are not suitable for everyday use and suffer from temporal delays of several seconds, they have good spatial resolution and, most importantly, can sample the activity of deep brain structures. Recently, fMRI was used to measure brain activation during the operation of a BCI based on slow cortical potentials [4].

Myoelectric systems that make use of voluntary activations of unaffected muscles in partially paralyzed subjects and amputees [36–39], and use these signals to control limb prostheses and exoskeletons, present an alternative to the existing non-invasive BCIs. Currently, these systems are more practical for everyday situations than EEG-based BCIs [11]. The details of their operation are beyond the scope of this review.

In summary, severely and partially paralyzed patients can reacquire basic forms of communication and motor control using EEG-based systems. Yet motor recovery obtained using these systems has been limited, and no clear breakthrough that could significantly enhance the power of EEG-based BCIs in the near future has been reported in the literature [11]. This by no means reduces the clinical utility of such systems. Some of them have improved the quality of life of patients, such as the BCI for spelling [3]. But if the goal of a BMI is to restore movements with multiple degrees of freedom through the control of an artificial prosthesis, the message from published evidence is clear: this task will require recording of high-resolution signals from the brain, and this can be done using invasive approaches.

Invasive BMIs

Invasive BMI approaches are based on recordings from ensembles of single brain cells (also known as single units) or on the activity of multiple neurons (also known as multi-units). These approaches have their roots in the pioneering studies conducted by Fetz and colleagues in the 1960s and 1970s [40–45]. In these experiments, monkeys learned to control the activity of their cortical neurons voluntarily, aided by biofeedback indicating the firing rate of single neurons. A few years after these experiments, Edward Schmidt raised the possibility that voluntary motor commands could be extracted from raw cortical neural activity and used to control a prosthetic device designed to restore motor functions in severely paralyzed patients [46].

Largely owing to technical difficulties associated with obtaining the needed cortical signals and implementing real-time interfaces quickly enough, thorough experimental testing of Schmidt's proposition took almost two decades to be accomplished. These bottlenecks were passed because of a series of experimental and technological breakthroughs that led to a new electrophysiological methodology for chronic, multi-site, multi-electrode recordings [47–51]. The BMI approach that relies on long-term recordings from large populations of neurons (100–400 units) evolved from experiments carried out in 1995 [47]. After the introduction of such an approach, a series of studies demonstrated that neuronal readout of tactile stimuli could be uncovered using pattern-recognition algorithms, such as artificial neural networks [52,53]. These developments paved the way for the first experiment in which neuronal population activity recorded in behaving rats enacted movements of a robotic device that had a single degree of freedom [1]. Soon after this first demonstration, a similar BMI approach was shown to work in New World [54] and rhesus monkeys [55–58]. As a result of these experimental efforts, in less than six years several laboratories reported BMIs that reproduced primate arm reaching [1,54–58] and the combination of reaching and grasping movements [57], using either computer cursors or robotic manipulators as actuators.

During the past three years, most of the published studies on BMIs have been conducted in behaving rhesus monkeys. There are several important differences that distinguish these BMIs (Figure 1). These include:

the number of cortical implants (e.g. uni-site or multi-site recordings); the cortical location of implants (e.g. frontal or parietal cortex, or both); the type of neural signal recorded (local field potentials versus single-unit or multi-unit signals); and the size of the neural sample. With the exception of the BMIs used at Duke University (<http://www.duke.edu/>), all BMIs tested in monkeys have relied on single cortical site recordings either of local field potentials [59–62] or of small samples (<30) of neurons or multi-units [55,56,63]. Most of these small-sample, single-area BMIs utilized neural signals recorded in the primary motor cortex [55,56], although one group has focused on BMIs that processed neural signals recorded in the posterior parietal cortex [64]. At Duke University, a BMI strategy has recently been implemented based on single-unit recordings made during intra-operative placement of deep-brain stimulators in Parkinsonian patients [65].

Principles of BMI operation

Invasive BMIs rely on the physiological properties of individual cortical and subcortical neurons (or pools of neurons) that modulate their activity in association with movements. First documented four decades ago by Evarts [66–68], such modulations are highly variable, from neuron to neuron and from trial to trial [69–72]. Thus, as much as neighboring neurons might display highly distinct firing modulation patterns during the execution of a particular movement, single-neuron firing can vary substantially from one trial to the next, despite the fact that the overt movements remain virtually identical. Yet averaging across many trials reveals fairly consistent firing patterns. By the same token, averaging across large populations of neurons significantly reduces the variability of signals derived from single neurons [54,69].

Extracting motor control signals from the firing patterns of populations of neurons and using these control signals to reproduce motor behaviors in artificial actuators are the two key operations that a clinically viable BMI should perform flawlessly [51,73]. To be accepted by patients, BMI devices will also have to act in the same way and feel the same as the subjects' own limbs. Recent findings suggest that this task might be accomplished by creating conditions under which the brain undergoes experience-dependent plasticity and assimilates the prosthetic limb as if it were part of the subject's own body. Until recently, such plasticity was achieved using visual feedback. However, a more efficient way to assimilate the prosthetic limb in the brain representation could be to use multiple artificial feedback signals, derived from pressure and position sensors placed on the prosthetic limb. These feedback signals would effectively train the brain to incorporate the properties of the artificial limb into the tuning characteristic of neurons located in cortical and subcortical areas that maintain representations of the subject's body. We predict that such plasticity will result in sensory and motor areas of the brain representing the prosthetic device.

A proposed roadmap for the future of BMI research

To achieve the ambitious goal of creating a clinically useful invasive BMI for restoring upper-limb mobility, one has to pass the following key bottlenecks:

- Obtaining stable, very long-term recordings (i.e. over years) of large populations of neurons (i.e. hundreds to thousands) from multiple brain areas. This task encourages development of a new generation of biocompatible 3D electrode matrices that yield thousands of channels of recordings while producing little tissue damage at implantation and minimal inflammatory reaction thereafter.
- Developing computationally efficient algorithms, that can be incorporated into the BMI software, for translating neuronal activity into high-precision command signals capable of controlling an artificial actuator that has multiple degrees of freedom.
- Learning how to use brain plasticity to incorporate prosthetic devices into the body representation. This will make the prosthetic feel like the subject's own limb.
- Implementing a new generation of upper-limb prosthetics, capable of accepting brain-derived control signals to perform movements with multiple degrees of freedom.

We now discuss some potential avenues for addressing the first three of these major challenges. A thorough discussion of the fourth challenge (i.e. engineering a new generation of prosthetic arms) is beyond the scope of this review.

Long-term recordings of neuronal activity from multiple brain areas

Although recording from single neurons is the first choice of neurophysiologists, multi-unit signals that comprise activity of a few neurons can also be efficiently used in BMI control [57]. In addition, several reports have suggested using local field potentials [59–62]. It is conceivable that, in future, advanced neuroprosthetic devices will use hybrid solutions in which a combination of several types of neural signals are recorded and processed. Here, however, we focus on using single-unit and multi-unit signals as the primary input to a BMI. This choice raises a fundamental question: how many neurons does a BMI need to sample to produce effective motor outputs?

This question, first raised several years ago [51,73], remains a matter of debate. Some groups [55,56,63] have strongly claimed that recordings from a small number of neurons can be sufficient for good performance of a BMI. Selected populations of highly tuned neurons can indeed accurately predict movement parameters [74]. However, highly tuned neurons are rare in a typical random sample of cortical cells. Given that the neuronal yield of all chronic recording techniques is produced by random sampling of neurons, it is unrealistic to expect that a large fraction of these cells will be highly tuned to a particular motor variable. Moreover, it would be even more unrealistic to expect that a small neuronal sample would represent several variables of interest. Therefore, large samples of recorded neurons are preferable, at the very least to enable selection of a sufficient number of highly tuned neurons. Besides, the reason for relying on large neuronal populations goes far beyond the issue of selecting the best performing cells. Both the accuracy [54,57,70] and the

reliability [69] of predictions improve considerably with the number of simultaneously recorded neurons, because motor information seems to be represented in the cortex in a highly distributed way. Thus, as the neuronal sample increases in size, errors related to individual neuron firing variability decrease significantly [69]. So looking into the future, it seems unlikely that invasive BMIs based on a small group of neurons will be capable of continuously reproducing in artificial limbs the range of fine movements normally performed by the human arm and hand.

Currently, chronically implanted microwire arrays offer the best compromise between safety, recording longevity and neuronal yield required to operate BMIs [48,51,54,57,58,73]. It is clear that this methodology will continue to be applied in experimental settings, but several significant improvements are required before it becomes fully applicable for long-term (months to years) chronic clinical applications in humans. First and foremost, the broad and challenging issue of biological compatibility [75–80] has to be properly addressed and solved. Second, fully implantable technologies using wireless headstages for amplification of neuronal signals have to be implemented to reduce the risks of infection introduced by the use of cables that connect brain implants to external hardware.

Current microelectrode designs typically enable good quality recordings to be made for several months. In certain cases and species, these recordings can last for several years [81]. However, recording quality often deteriorates, probably owing to a process of electrode encapsulation by fibrous tissue and cell death in the vicinity of the electrode [77]. Some authors have proposed that electrodes that contain neurotrophic medium [82–85], or are coated with factors that promote neuronal growth (e.g. nerve growth factor, brain-derived neurotrophic factor or laminin) and various anti-inflammatory compounds (e.g. dexamethazone) [77,86–91], might be a way to cope with encapsulation. Currently, it is unclear whether these approaches will be useful.

Efforts to resolve the biocompatibility issues will probably have to be pursued in parallel with the development of new 3D electrode matrices, which should aim to increase the average yield to thousands of neuronal signals per implanted probe. Current alternatives to such microwires (e.g. the Utah probe [92], which implements arrays of rigid, single-ended electrodes) have yet to prove their adequacy to support the haste in which this technology was moved into clinical applications. Judging from the published evidence, such arrays are best suited to sample neuronal activity from flat surfaces of cortical gyri in animal experiments. However, this design might not be suitable for long-term use in human patients. In addition to issues of how electrodes are inserted into the cortex, the inability to sample from deep cortical layers, and many unanswered biocompatibility questions, the reliability of the recording system utilized by currently available probes is also compromised by the continuous stress of a daily routine that involves external cables and plugging and unplugging of external head-stages. These operations carry a risk of causing tissue damage, bleeding and brain infection. Such a risk of failure, which might be tolerable in animal experiments, is unwanted in practical applications

for humans. Cyberkinetics Neurotechnology Systems (<http://www.cyberkineticsinc.com/content/index.jsp>) has recently started clinical trials in severely paralyzed patients of a BCI based on a probe developed at the University of Utah (<http://www.utah.edu/>). Because no peer-reviewed publication has appeared related to this work, the exact outcome of this study remains unknown.

From these considerations, it is clear that the issues related to the long-term functionality of implantable micro-electrodes, and to the development of fully implantable electronic devices for amplification of a large number of neuronal signals and their wireless transmission to the actuator, are the major technological challenges that will determine the success or failure of future clinical applications of BMI technology. These technological developments are necessary not only to increase the practical usefulness of the BMI (more neurons mean better stability and accuracy) but also to ensure that risks to patient health are minimized. Auspiciously, telemetry transmission methods [93–97] for effective wireless transmission of multi-channel neuronal signals have already started to appear in the literature [98,99]. These solutions are currently being tested in animal experiments.

Many new ideas of how to improve neuronal recordings have been proposed recently. These range from ceramic-based multi-electrode arrays [100] to nanotechnology probes that access the brain through the vascular system [101]. In this latter design, probes record neuronal activity without compromising brain parenchyma. Undoubtedly, much more testing will be needed to conclude which of these ideas are viable.

Developing algorithms for translating neuronal activity into command signals for artificial actuators

Currently, neuroscientists are far from obtaining a clear understanding of how motor and cognitive information is processed by the populations of neurons that form large brain circuits. Rate encoding, temporal encoding and population encoding principles have been suggested, and various experimental paradigms, including BMIs, have been developed to test the validity of these concepts. However, precise knowledge of computations performed by brain circuits is not crucial for the construction of clinically relevant BMIs. Mostly, BMI platforms take advantage of the well known correlation between discharges of cortical neurons and motor parameters of interest, and perform a reverse operation: they predict motor parameters from patterns of neuronal firing. Generally, predictions of motor parameters do not signify a causal relationship between the neuronal activity and the generation of movements. One type of correlation between neuronal activity and movement is known as directional tuning [102,103], and correlations of neuronal activity with kinematic [104–106] and kinetic [107–109] parameters of movements have also been described.

Although a wealth of linear and nonlinear algorithms for translating neuronal activity into commands to artificial actuators have been suggested [1,54,56,57,70,110–116], relatively simple multiple linear regression models have proved to be efficient in many practical BMI designs [54,55,57,58,65,117]. In these models, predicted motor

parameters are derived from the weighted sums of neuronal rates of firing, measured at different time points in the past. The number of regressors in the model and the time window used for predictions can be optimized for each concrete BMI application [70,74,117]. Linear methods, such as population vector predictions, can incorporate adaptive algorithms that continuously update the model parameters while the subject trains [56].

Basic research using the BMI paradigm has lent support to some fundamental principles of neural information coding. For example, studies in which several independent linear models were run in parallel revealed that several motor parameters, such as arm position, velocity, acceleration and hand gripping force, could be predicted simultaneously by separate combinations of the activity of the same original pool of neurons [57]. This finding supports the notion that multiple motor parameters are processed by overlapping neuronal ensembles. As a corollary, a single cortical neuron can contribute to several predictions simultaneously.

The choice of motor parameters extracted in future clinical BMIs will depend on the main rehabilitation or therapeutic goals of these applications. For example, an experimental BMI for reaching and grasping [57] predicted hand velocity and gripping force that matched the characteristics of a robotic arm equipped with a gripper. In the near future, this design could lead to the implementation of neuroprosthetic devices that help quadriplegic or 'locked in' patients to reach and grasp objects in the surrounding space. We can also envisage that BMIs that synthesize speech, based on neuronal signals recorded in intact speech-related regions, could one day help patients suffering from aphasia due to cortical strokes recover their ability to communicate.

Future clinical applications might also take advantage of BMIs that predict EMG signals [117]. The main benefit of this design compared with a BMI that predicts kinematic parameters is that the signals of individual muscles can control biologically-inspired devices, which would produce a whole range of actuator stiffness. This is an important property needed for a future generation of prosthetic limbs that should be able to manipulate objects with different physical characteristics. Another powerful future application for BMIs that decode EMGs is the construction of brain-muscle interfaces that directly stimulate the muscles of paralyzed patients and thereby restore mobility by using the patient's own musculoskeletal apparatus [37,118,119]. Such BMIs are likely to be much more acceptable to patients, particularly because the hardware needed for amplification, transmission and processing of brain-derived control signals, and the muscle stimulators driven by these neural signals, such as the BION [120], can be entirely encased in the patient's body. In the future, it is conceivable that such BMIs could merge the current cortically driven paradigm with methods and new technologies developed in the field of functional electrical stimulation.

Early BMI designs focused on decoding motor parameters from neuronal activity [1,54–56,117]. More recently, it was suggested that BMIs that decode cognitive signals, for example those that decode intended reach direction during the delay periods preceding movement

execution [64,121–123], could also be efficient. Although this idea is very attractive, a BMI based exclusively on cognitive signals cannot execute continuous control of movement parameters. Instead, it decodes higher-order characteristics of movements, such as reach direction or characteristics of objects being grasped, and delegates lower-order details of motor execution to the actuator controller. Recently, we have proposed that a hybrid BMI, based on a shared control mode of operation [124], can improve the accuracy with which the prosthetic implements the voluntary intentions of the subject. A shared-control mode of operation would be achieved by a combination of high-order brain-derived signals, conveying the subject's voluntary intentions, and low-level artificial 'reflex-like' circuits, designed to improve the precision with which prosthetic limb movements are generated.

In the future, BMIs that take advantage of the higher-order neuronal representations of movement-related variables will also emerge. These BMIs will probably derive information from representations of movement sequences [125,126], reference frames [127–130], potential movement targets [131] and simultaneous encoding of multiple spatial variables, such as movement direction, orientation of selective spatial attention, and gaze angle [132,133]. Utilization of this wide array of information will endow BMIs with a much more flexible control of prosthetic limbs.

In the same context, we also believe that future BMIs will take advantage of new insights on how neural circuits encode temporal characteristics of movements. Recent studies [134–137] have indicated that a rather distributed representation of temporal information might exist in the brain. Recordings obtained from primary motor and premotor cortical ensembles while monkeys performed self-timed button presses [138] enabled prediction of both the time that had elapsed since the monkey pressed the button and the time until the button would be released. Because any motor behavior has a temporal structure, and because episodes of movement execution are typically intermingled with periods of immobility during which movements are being prepared, a BMI that decodes behavioral time will be able to inhibit movements of the actuator during waiting periods and release the actuator at appropriate times. These operations will be based on the voluntary intentions of the user.

Making use of brain plasticity to incorporate prosthetic devices into the body representation

Controlling an artificial actuator through a BMI can be thought of as a process somewhat similar to the operation required by subjects to operate tools – a capacity that is inherent only in higher primates such as chimpanzees and humans [139]. Almost 100 years ago [140], Head and Holmes suggested that the 'body schema' – that is, the internal brain representations of one's body – could extend itself to include a wielded tool. This idea was validated by the experimental demonstration that cortical neurons extend their visual receptive fields along the length of a rake used by monkeys to retrieve distant objects [141]. Psychophysics experiments also support the notion that tool usage leads to remapping of the 'body schema' in

humans [142,143]. Accordingly, a recent neuroimaging study [144] described specific activations of the right ventral premotor cortex during manipulation of a myoelectric prosthetic hand. Altogether, these results suggest that long-term usage of an artificial actuator directly controlled by brain activity might lead to substantial cortical and subcortical remapping. As such, this process might elicit the vivid perceptual experience that the artificial actuator becomes an extension of the subject's body rather than a mere tool. This suggestion is supported by the report of primary sensorimotor cortex activation during perceived voluntary movements of phantom limbs in amputees [145].

Perhaps the most stunning demonstration of tool assimilation by animals was observed when both rats and primates learned to operate an actuator through a BMI, without the need to move their own limbs [1,56–58]. In these experiments, decoding algorithms were initially trained to predict limb movements of animals from the activity of neuronal populations. Remarkably, after these animals started to control the actuator directly using their neuronal activity, their limbs eventually stopped moving, while the animals continued to control the actuator by generating proper modulations of their cortical neurons. Interestingly, during these episodes neuronal tuning to movements of the subject's own limbs decreased while the animals continued to control the artificial actuator by their brain activity [58]. The most parsimonious interpretation of this finding is that the brain was capable of undergoing a gradual assimilation of the actuator within the same maps that represented the body [57,58]. Neuronal mechanisms mediating such plasticity are far from being understood.

However, it is fair to state that there is a growing consensus in the literature that continuous BMI operations in primates lead to physiological changes in neuronal tuning, which include changes in preferred direction and direction tuning strength of neurons [56–58]. In addition, broad changes in pair-wise neuronal correlation can be detected after BMIs are switched to operate fully under brain-control mode [57,58].

Along with these physiological adaptations of neuronal firing patterns, behavioral performance improves as animals learn to operate BMIs effectively [56–58]. Initial training to operate a BMI is characterized by an increase in neuronal firing rate variance, which cannot be simply explained by changes in limb or actuator movements [146]. As the quality of BMI control improves, initial elevation of neuronal firing variability subsides. Plastic changes in neuronal firing patterns during BMI control, leading to the physiological incorporation of the artificial actuator properties into neuronal space, could account for these changes in firing rate variance. This interpretation is in accord with the theory of optimal feedback control [147–149]. According to this theory, a motor system acts as a stochastic feedback controller that optimizes only those motor parameters that are necessary to achieve the goals of a particular task. During the brain-control mode of operation of a BMI, the goals of a motor task are achieved only by direct brain control of an artificial actuator. Thus, in terms of optimal feedback control theory, neuronal ensembles should adapt their physiological tuning properties to represent better the goal-related variables of the task performed by the BMI.

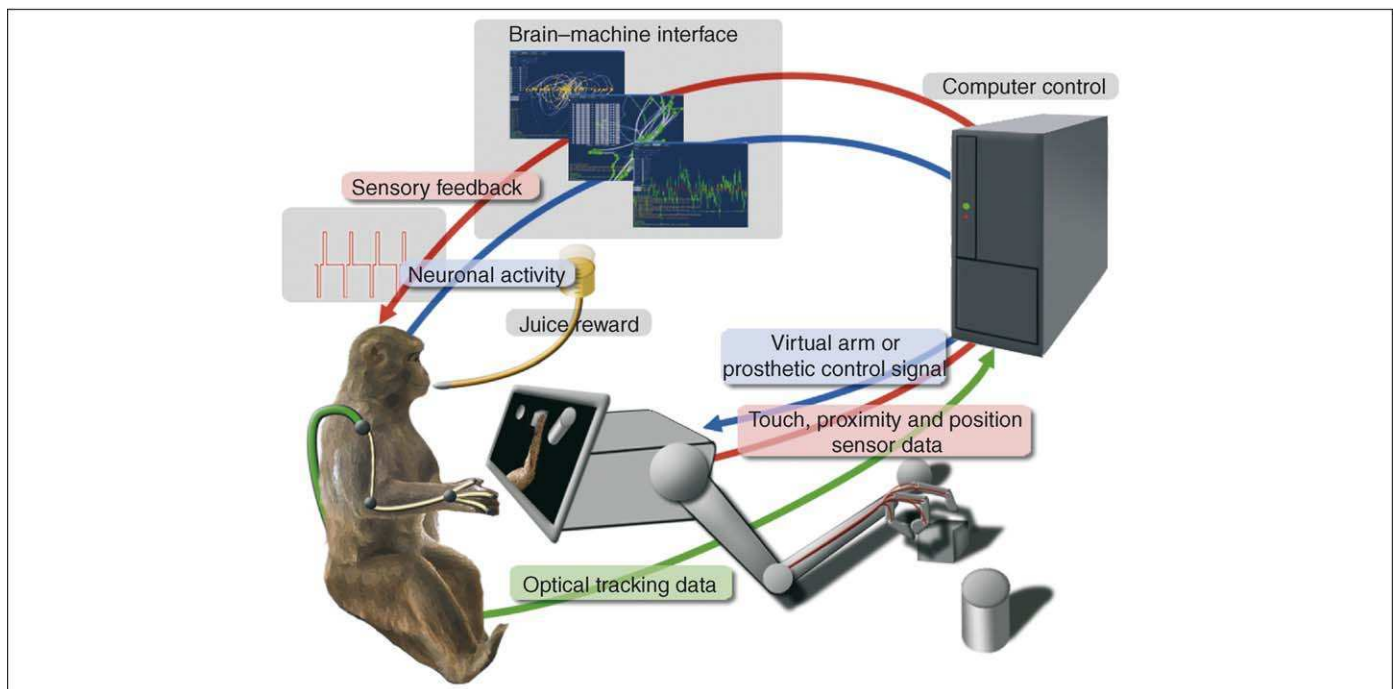


Figure 2. A BMI with multiple feedback loops being developed at the Duke University Center for Neuroengineering. A rhesus macaque is operating an artificial robotic manipulator that reaches and grasps different objects. The manipulator is equipped with touch, proximity and position sensors. Signals from the sensors are delivered to the control computer (right), which processes them and converts to microstimulation pulses delivered to the sensory areas in the brain of the monkey, to provide it with feedback information (red loop). A series of microstimulation pulses is illustrated in the inset on the left. Neuronal activity is recorded in multiple brain areas and translated to commands to the actuator, via the control computer and multiple decoding algorithms (blue loop). Arm position is monitored using an optical tracking system that tracks the position of several markers mounted on the arm (green loop). We hypothesize that continuous operation of this interface would lead to incorporation of the external actuator into the representation of the body in the brain. Figure designed by Nathan Fitzsimmons.

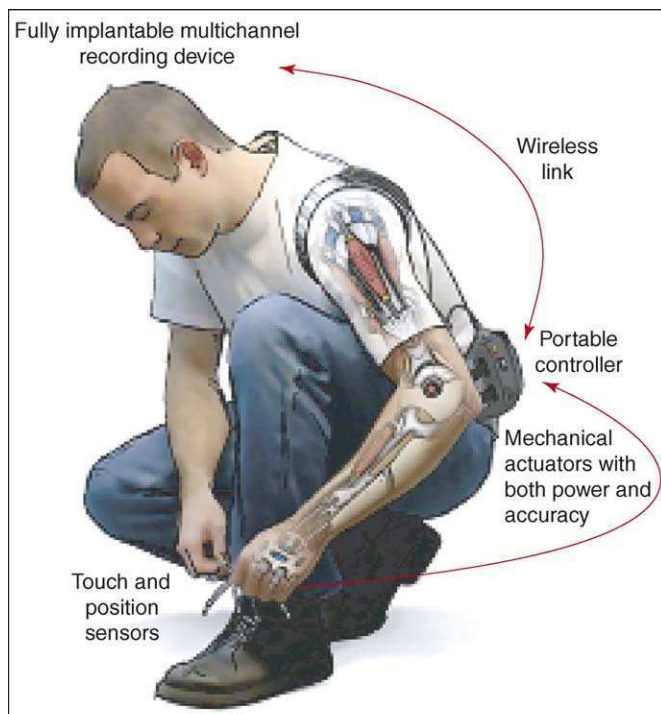


Figure 3. How a fully-implantable BMI could restore limb mobility in paralyzed subjects or amputees. Although the details of this system have to be worked out through future research, it is clear that the BMI for human clinical applications should be encased in the patient's body as much as possible. Wireless telemetry offers a viable solution for this purpose. The prosthesis not only should have the functionality of the human arm in terms of power and accuracy of the actuators, but also should be equipped with the sensors of touch and position from which signals can be transmitted back to the subject's brain.

Making the prosthetic feel like the subject's own limb using microstimulation of cortical sensory areas

Peripheral tactile and proprioceptive signals contribute to the normal operation of one's limbs and the perception that they are part of the body [142,143]. For a neuroprosthesis to behave and feel as a natural appendage of the subject's body, it will have to be instrumented with various sensors that can provide multiple channels of 'sensory' information back to the subject's brain. In most current BMI designs, animal subjects receive sensory information from the actuator through visual feedback [55–58]. Predictions of motor parameters are less stable in the absence of visual feedback [70] than when it is present [55–58].

Curiously, the use of tactile and proprioceptive-like feedback in BMI research remains largely unexplored. Recently, in collaboration with John Chapin, we have started to explore the intriguing possibility of delivering such sensory feedback information, generated in the actuator, to the brain through the use of multi-channel microstimulation of somatosensory cortical areas (Figure 2). Previous studies have shown that monkeys sense microstimulation patterns and can use them to guide their behavioral responses [150,151]. In a recent long-term study, owl monkeys could learn to guide their reaching movements by decoding vibratory stimuli applied to their arms [152]. Next, instead of vibratory stimulation, matching patterns of microstimulation were applied through the electrodes implanted in the primary somatosensory cortex [153]. Monkeys were still able to interpret correctly the

instructions provided by cortical microstimulation, and their behavioral performance eventually surpassed the level of performance observed when the vibratory stimulus was applied to their skin. These results suggest that cortical microstimulation might become a useful way to deliver long-term feedback from prosthetic limbs controlled by a BMI, and might contribute to the development of a completely new generation of neuroprosthetic devices for restoring various motor behaviors in severely impaired patients.

Concluding remarks

Our vision of neuroprosthetic developments that might emerge in the next 10–20 years includes a fully implantable recording system that wirelessly transmits multiple streams of electrical signals, derived from thousands of neurons, to a BMI capable of decoding spatial and temporal characteristics of movements and intermittent periods of immobility, in addition to cognitive characteristics of the intended actions (Figure 3). This BMI would utilize a combination of high-order motor commands, derived from cortical and subcortical neuronal activity, and peripheral low-level control signals, derived from artificial 'reflex-like' control loops. Such a shared-control mode of BMI operation would either command an actuator with multiple degrees of freedom or directly stimulate multiple peripheral nerves and muscles through implantable stimulators. Highly instrumented artificial actuators, containing arrays of touch and position sensors, would generate multiple streams of sensory feedback signals that could be directly delivered to cortical and/or subcortical somatosensory areas of the subject's brain, through spatiotemporal patterns of multi-channel microstimulation. Such closed-loop, hybrid BMIs would get one step closer to the dream of restoring a large repertoire of motor functions to a multitude of patients who currently have very few options for regaining their mobility.

References

- Chapin, J.K. *et al.* (1999) Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nat. Neurosci.* 2, 664–670
- Wolpaw, J.R. *et al.* (2002) Brain–computer interfaces for communication and control. *Clin. Neurophysiol.* 113, 767–791
- Birbaumer, N. *et al.* (1999) A spelling device for the paralysed. *Nature* 398, 297–298
- Hinterberger, T. *et al.* (2005) Neuronal mechanisms underlying control of a brain–computer interface. *Eur. J. Neurosci.* 21, 3169–3181
- Kubler, A. *et al.* (2001) Brain–computer communication: unlocking the locked in. *Psychol. Bull.* 127, 358–375
- Kubler, A. *et al.* (2001) Brain–computer communication: self-regulation of slow cortical potentials for verbal communication. *Arch. Phys. Med. Rehabil.* 82, 1533–1539
- Obermaier, B. *et al.* (2003) Virtual keyboard controlled by spontaneous EEG activity. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 422–426
- Obermaier, B. *et al.* (2001) Information transfer rate in a five-classes brain–computer interface. *IEEE Trans. Neural Syst. Rehabil. Eng.* 9, 283–288
- Sheikh, H. *et al.* (2003) Electroencephalographic (EEG)-based communication: EEG control versus system performance in humans. *Neurosci. Lett.* 345, 89–92
- Wolpaw, J.R. (2004) Brain–computer interfaces (BCIs) for communication and control: a mini-review. *Suppl. Clin. Neurophysiol.* 57, 607–613

- 11 Birbaumer, N. (2006) Brain-computer-interface research: coming of age. *Clin. Neurophysiol.* 117, 479–483
- 12 Nowlis, D.P. and Kamiya, J. (1970) The control of electroencephalographic alpha rhythms through auditory feedback and the associated mental activity. *Psychophysiology* 6, 476–484
- 13 Plotkin, W.B. (1976) On the self-regulation of the occipital alpha rhythm: control strategies, states of consciousness, and the role of physiological feedback. *J. Exp. Psychol. Gen.* 105, 66–99
- 14 Wyricka, W. and Serman, M. (1968) Instrumental conditioning of sensorimotor cortex EEG spindles in the waking cat. *Psychol Behav* 3, 703–707
- 15 Serman, M.B. *et al.* (1974) Biofeedback training of the sensorimotor electroencephalogram rhythm in man: effects on epilepsy. *Epilepsia* 15, 395–416
- 16 Black, A.H. (1971) The direct control of neural processes by reward and punishment. *Am. Sci.* 59, 236–245
- 17 Middendorf, M. *et al.* (2000) Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Trans. Rehabil. Eng.* 8, 211–214
- 18 Sutter, E.E. and Tran, D. (1992) The field topography of ERG components in man – I. The photopic luminance response. *Vision Res.* 32, 433–446
- 19 Kelly, S.P. *et al.* (2005) Visual spatial attention control in an independent brain-computer interface. *IEEE Trans. Biomed. Eng.* 52, 1588–1596
- 20 Donchin, E. *et al.* (2000) The mental prosthesis: assessing the speed of a P300-based brain-computer interface. *IEEE Trans. Rehabil. Eng.* 8, 174–179
- 21 Piccione, F. *et al.* (2006) P300-based brain computer interface: reliability and performance in healthy and paralysed participants. *Clin. Neurophysiol.* 117, 531–537
- 22 Sellers, E.W. and Donchin, E. (2006) A P300-based brain-computer interface: initial tests by ALS patients. *Clin. Neurophysiol.* 117, 538–548
- 23 Birbaumer, N. *et al.* (2000) The thought translation device (TTD) for completely paralyzed patients. *IEEE Trans. Rehabil. Eng.* 8, 190–193
- 24 Pfurtscheller, G. *et al.* (2006) Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. *NeuroImage* 33, 153–159
- 25 Pfurtscheller, G. *et al.* (2003) Graz-BCI: state of the art and clinical applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* 11, 177–180
- 26 Wolpaw, J.R. and McFarland, D.J. (2004) Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17849–17854
- 27 Pfurtscheller, G. and Lopes da Silva, F.H. (1999) Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.* 110, 1842–1857
- 28 Bayliss, J.D. and Ballard, D.H. (2000) A virtual reality testbed for brain-computer interface research. *IEEE Trans. Rehabil. Eng.* 8, 188–190
- 29 Pfurtscheller, G. *et al.* (2006) Walking from thought. *Brain Res.* 1071, 145–152
- 30 Leuthardt, E.C. *et al.* (2004) A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* 1, 63–71
- 31 Hinterberger, T. *et al.* (2003) A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. *Clin. Neurophysiol.* 114, 416–425
- 32 Kubler, A. *et al.* (2005) Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology* 64, 1775–1777
- 33 Pfurtscheller, G. *et al.* (2003) ‘Thought’ control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia. *Neurosci. Lett.* 351, 33–36
- 34 Keith, M.W. *et al.* (1989) Implantable functional neuromuscular stimulation in the tetraplegic hand. *J. Hand Surg. (Am.)* 14, 524–530
- 35 Weiskopf, N. *et al.* (2004) Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *J. Physiol. (Paris)* 98, 357–373
- 36 Light, C.M. *et al.* (2002) Intelligent multifunction myoelectric control of hand prostheses. *J. Med. Eng. Technol.* 26, 139–146
- 37 Navarro, X. *et al.* (2005) A critical review of interfaces with the peripheral nervous system for the control of neuroprostheses and hybrid bionic systems. *J. Peripher. Nerv. Syst.* 10, 229–258
- 38 Okuno, R. *et al.* (2005) Compliant grasp in a myoelectric hand prosthesis. Controlling flexion angle and compliance with electromyogram signals. *IEEE Eng. Med. Biol. Mag.* 24, 48–56
- 39 Zecca, M. *et al.* (2002) Control of multifunctional prosthetic hands by processing the electromyographic signal. *Crit. Rev. Biomed. Eng.* 30, 459–485
- 40 Fetz, E.E. (1969) Operant conditioning of cortical unit activity. *Science* 163, 955–958
- 41 Fetz, E.E. (1992) Are movement parameters recognizably coded in activity of single neurons? *Behav Brain Sci.* 15, 679–690
- 42 Fetz, E.E. and Baker, M.A. (1973) Operantly conditioned patterns on precentral unit activity and correlated responses in adjacent cells and contralateral muscles. *J. Neurophysiol.* 36, 179–204
- 43 Fetz, E.E. and Finocchio, D.V. (1971) Operant conditioning of specific patterns of neural and muscular activity. *Science* 174, 431–435
- 44 Fetz, E.E. and Finocchio, D.V. (1972) Operant conditioning of isolated activity in specific muscles and precentral cells. *Brain Res.* 40, 19–23
- 45 Fetz, E.E. and Finocchio, D.V. (1975) Correlations between activity of motor cortex cells and arm muscles during operantly conditioned response patterns. *Exp. Brain Res.* 23, 217–240
- 46 Schmidt, E.M. (1980) Single neuron recording from motor cortex as a possible source of signals for control of external devices. *Ann. Biomed. Eng.* 8, 339–349
- 47 Nicolelis, M.A. *et al.* (1995) Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268, 1353–1358
- 48 Nicolelis, M.A. *et al.* (2003) Chronic, multisite, multielectrode recordings in macaque monkeys. *Proc. Natl. Acad. Sci. U. S. A.* 100, 11041–11046
- 49 Nicolelis, M.A. *et al.* (1997) Reconstructing the engram: simultaneous, multisite, many single neuron recordings. *Neuron* 18, 529–537
- 50 Nicolelis, M.A. and Ribeiro, S. (2002) Multielectrode recordings: the next steps. *Curr. Opin. Neurobiol.* 12, 602–606
- 51 Nicolelis, M.A. (2001) Actions from thoughts. *Nature* 409, 403–407
- 52 Ghazanfar, A.A. *et al.* (2000) Encoding of tactile stimulus location by somatosensory thalamocortical ensembles. *J. Neurosci.* 20, 3761–3775
- 53 Krupa, D.J. *et al.* (2004) Layer-specific somatosensory cortical activation during active tactile discrimination. *Science* 304, 1989–1992
- 54 Wessberg, J. *et al.* (2000) Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 408, 361–365
- 55 Serruya, M.D. *et al.* (2002) Instant neural control of a movement signal. *Nature* 416, 141–142
- 56 Taylor, D.M. *et al.* (2002) Direct cortical control of 3D neuroprosthetic devices. *Science* 296, 1829–1832
- 57 Carmena, J.M. *et al.* (2003) Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* 1, E42
- 58 Lebedev, M.A. *et al.* (2005) Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *J. Neurosci.* 25, 4681–4693
- 59 Mehring, C. *et al.* (2003) Inference of hand movements from local field potentials in monkey motor cortex. *Nat. Neurosci.* 6, 1253–1254
- 60 Rickert, J. *et al.* (2005) Encoding of movement direction in different frequency ranges of motor cortical local field potentials. *J. Neurosci.* 25, 8815–8824
- 61 Pesaran, B. *et al.* (2002) Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat. Neurosci.* 5, 805–811
- 62 Scherberger, H. *et al.* (2005) Cortical local field potential encodes movement intentions in the posterior parietal cortex. *Neuron* 46, 347–354
- 63 Tillery, S.I. and Taylor, D.M. (2004) Signal acquisition and analysis for cortical control of neuroprosthetics. *Curr. Opin. Neurobiol.* 14, 758–762
- 64 Musallam, S. *et al.* (2004) Cognitive control signals for neural prosthetics. *Science* 305, 258–262
- 65 Patil, P.G. *et al.* (2004) Ensemble recordings of human subcortical neurons as a source of motor control signals for a brain-machine interface. *Neurosurgery* 55, 27–35

- 66 Evarts, E.V. (1966) Pyramidal tract activity associated with a conditioned hand movement in the monkey. *J. Neurophysiol.* 29, 1011–1027
- 67 Evarts, E.V. (1968) Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* 31, 14–27
- 68 Evarts, E.V. (1968) A technique for recording activity of subcortical neurons in moving animals. *Electroencephalogr. Clin. Neurophysiol.* 24, 83–86
- 69 Carmenta, J.M. *et al.* (2005) Stable ensemble performance with single-neuron variability during reaching movements in primates. *J. Neurosci.* 25, 10712–10716
- 70 Wessberg, J. and Nicolelis, M.A. (2004) Optimizing a linear algorithm for real-time robotic control using chronic cortical ensemble recordings in monkeys. *J. Cogn. Neurosci.* 16, 1022–1035
- 71 Stein, R.B. *et al.* (2005) Neuronal variability: noise or part of the signal? *Nat. Rev. Neurosci.* 6, 389–397
- 72 Cohen, D. and Nicolelis, M.A. (2004) Reduction of single-neuron firing uncertainty by cortical ensembles during motor skill learning. *J. Neurosci.* 24, 3574–3582
- 73 Nicolelis, M.A. (2003) Brain–machine interfaces to restore motor function and probe neural circuits. *Nat. Rev. Neurosci.* 4, 417–422
- 74 Sanchez, J.C. *et al.* (2004) Ascertaining the importance of neurons to develop better brain–machine interfaces. *IEEE Trans. Biomed. Eng.* 51, 943–953
- 75 Dodson, R.F. *et al.* (1978) Cerebral tissue response to electrode implantation. *Can. J. Neurol. Sci.* 5, 443–446
- 76 Schultz, R.L. and Willey, T.J. (1976) The ultrastructure of the sheath around chronically implanted electrodes in brain. *J. Neurocytol.* 5, 621–642
- 77 Polikov, V.S. *et al.* (2005) Response of brain tissue to chronically implanted neural electrodes. *J. Neurosci. Methods* 148, 1–18
- 78 Tresco, P.A. *et al.* (2000) Cellular transplants as sources for therapeutic agents. *Adv. Drug Deliv. Rev.* 42, 3–27
- 79 Berry, M. *et al.* (1999) *Cellular Responses to Penetrating CNS Injury*. CRC Press
- 80 Landis, D.M. (1994) The early reactions of non-neuronal cells to brain injury. *Annu. Rev. Neurosci.* 17, 133–151
- 81 Sandler, A.J. *et al.* (2005) Long-term neuronal recordings from nonhuman primates. In *2005 Abstract Viewer and Itinerary Planner*, Program No. 402.8, Society for Neuroscience Online (<http://sfn.scholarone.com/>)
- 82 Kennedy, P.R. (1989) The cone electrode: a long-term electrode that records from neurites grown onto its recording surface. *J. Neurosci. Methods* 29, 181–193
- 83 Kennedy, P.R. and Bakay, R.A. (1998) Restoration of neural output from a paralyzed patient by a direct brain connection. *NeuroReport* 9, 1707–1711
- 84 Kennedy, P.R. *et al.* (2000) Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* 8, 198–202
- 85 Kennedy, P.R. *et al.* (1992) The cone electrode: ultrastructural studies following long-term recording in rat and monkey cortex. *Neurosci. Lett.* 142, 89–94
- 86 Cui, X. *et al.* (2001) Surface modification of neural recording electrodes with conducting polymer/biomolecule blends. *J. Biomed. Mater. Res.* 56, 261–272
- 87 Rahimi, O. and Juliano, S.L. (2001) Transplants of NGF-secreting fibroblasts restore stimulus-evoked activity in barrel cortex of basal-forebrain-lesioned rats. *J. Neurophysiol.* 86, 2081–2096
- 88 Ignatius, M.J. *et al.* (1998) Bioactive surface coatings for nanoscale instruments: effects on CNS neurons. *J. Biomed. Mater. Res.* 40, 264–274
- 89 Kam, L. *et al.* (2002) Selective adhesion of astrocytes to surfaces modified with immobilized peptides. *Biomaterials* 23, 511–515
- 90 Cui, X. *et al.* (2003) *In vivo* studies of polypyrrole/peptide coated neural probes. *Biomaterials* 24, 777–787
- 91 Biran, R. *et al.* (2003) Directed nerve outgrowth is enhanced by engineered glial substrates. *Exp. Neurol.* 184, 141–152
- 92 Rousche, P.J. and Normann, R.A. (1998) Chronic recording capability of the Utah intracortical electrode array in cat sensory cortex. *J. Neurosci. Methods* 82, 1–15
- 93 Mohseni, P. *et al.* (2005) Wireless multichannel biopotential recording using an integrated FM telemetry circuit. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 263–271
- 94 Mackay, S. (1998) *Bio-Medical Telemetry: Sensing and Transmitting Biological Information from Animals and Man*. Wiley-IEEE Press
- 95 Knutti, J.W. *et al.* (1979) An integrated circuit approach to totally implantable telemetry systems. *Biotelem. Patient Monit.* 6, 95–106
- 96 Claude, J.P. *et al.* (1979) Applications of totally implantable telemetry systems to chronic medical research. *Biotelem. Patient Monit.* 6, 160–171
- 97 Chien, C.N. and Jaw, F.S. (2005) Miniature telemetry system for the recording of action and field potentials. *J. Neurosci. Methods* 147, 68–73
- 98 Bossetti, C.A. *et al.* (2004) Transmission latencies in a telemetry-linked brain–machine interface. *IEEE Trans. Biomed. Eng.* 51, 919–924
- 99 Morizio, J. *et al.* (2005) Fifteen-channel wireless headstage system for single-unit rat recordings. In *2005 Abstract Viewer and Itinerary Planner*, Program No. 68.4, Society for Neuroscience Online (<http://sfn.scholarone.com/>)
- 100 Moxon, K.A. *et al.* (2004) Ceramic-based multisite electrode arrays for chronic single-neuron recording. *IEEE Trans. Biomed. Eng.* 51, 647–656
- 101 Llinas, R.R. *et al.* (2005) Neuro-vascular central nervous recording/stimulating system: using nanotechnology probes. *J. Nanopart. Res.* 7, 111–127
- 102 Georgopoulos, A.P. *et al.* (1988) Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding of the direction of movement by a neuronal population. *J. Neurosci.* 8, 2928–2937
- 103 Georgopoulos, A.P. *et al.* (1986) Neuronal population coding of movement direction. *Science* 233, 1416–1419
- 104 Ashe, J. and Georgopoulos, A.P. (1994) Movement parameters and neural activity in motor cortex and area 5. *Cereb. Cortex* 4, 590–600
- 105 Moran, D.W. and Schwartz, A.B. (1999) Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* 82, 2676–2692
- 106 Averbeck, B.B. *et al.* (2005) Parietal representation of hand velocity in a copy task. *J. Neurophysiol.* 93, 508–518
- 107 Sergio, L.E. and Kalaska, J.F. (1998) Changes in the temporal pattern of primary motor cortex activity in a directional isometric force versus limb movement task. *J. Neurophysiol.* 80, 1577–1583
- 108 Sergio, L.E. *et al.* (2005) Motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks. *J. Neurophysiol.* 94, 2353–2378
- 109 Todorov, E. (2000) Direct cortical control of muscle activation in voluntary arm movements: a model. *Nat. Neurosci.* 3, 391–398
- 110 Truccolo, W. *et al.* (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol.* 93, 1074–1089
- 111 Brown, E.N. *et al.* (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461
- 112 Brockwell, A.E. *et al.* (2004) Recursive Bayesian decoding of motor cortical signals by particle filtering. *J. Neurophysiol.* 91, 1899–1907
- 113 Hu, J. *et al.* (2005) Feature detection in motor cortical spikes by principal component analysis. *IEEE Trans. Neural Syst. Rehabil. Eng.* 13, 256–262
- 114 Wu, W. *et al.* (2004) Modeling and decoding motor cortical activity using a switching Kalman filter. *IEEE Trans. Biomed. Eng.* 51, 933–942
- 115 Kim, S.P. *et al.* (2003) Divide-and-conquer approach for brain machine interfaces: nonlinear mixture of competitive linear models. *Neural Netw.* 16, 865–871
- 116 Kemere, C. *et al.* (2004) Model-based neural decoding of reaching movements: a maximum likelihood approach. *IEEE Trans. Biomed. Eng.* 51, 925–932
- 117 Santucci, D.M. *et al.* (2005) Frontal and parietal cortical ensembles predict single-trial muscle activity during reaching movements in primates. *Eur. J. Neurosci.* 22, 1529–1540
- 118 Degnan, G.G. *et al.* (2002) Functional electrical stimulation in tetraplegic patients to restore hand function. *J. Long Term Eff. Med. Implants* 12, 175–188

- 119 Peckham, P.H. and Knutson, J.S. (2005) Functional electrical stimulation for neuromuscular applications. *Annu. Rev. Biomed. Eng.* 7, 327–360
- 120 Loeb, G.E. *et al.* (2001) BION system for distributed neural prosthetic interfaces. *Med. Eng. Phys.* 23, 9–18
- 121 Hatsopoulos, N. *et al.* (2004) Decoding continuous and discrete motor behaviors using motor and premotor cortical ensembles. *J. Neurophysiol.* 92, 1165–1174
- 122 Rizzuto, D.S. *et al.* (2005) Spatial selectivity in human ventrolateral prefrontal cortex. *Nat. Neurosci.* 8, 415–417
- 123 Shenoy, K.V. *et al.* (2003) Neural prosthetic control signals from plan activity. *NeuroReport* 14, 591–596
- 124 Kim, H.K. *et al.* (2006) Continuous shared control stabilizes reach and grasping with brain-machine interfaces. *IEEE Trans. Biomed. Eng.* 53, 1164–1173
- 125 Hoshi, E. and Tanji, J. (2004) Differential roles of neuronal activity in the supplementary and presupplementary motor areas: from information retrieval to motor planning and execution. *J. Neurophysiol.* 92, 3482–3499
- 126 Lu, X. and Ashe, J. (2005) Anticipatory activity in primary motor cortex codes memorized movement sequences. *Neuron* 45, 967–973
- 127 Olson, C.R. (2003) Brain representation of object-centered space in monkeys and humans. *Annu. Rev. Neurosci.* 26, 331–354
- 128 Batista, A.P. *et al.* (1999) Reach plans in eye-centered coordinates. *Science* 285, 257–260
- 129 Battaglia-Mayer, A. *et al.* (2000) Early coding of reaching in the parietooccipital cortex. *J. Neurophysiol.* 83, 2374–2391
- 130 Graziano, M.S. and Gross, C.G. (1998) Spatial maps for the control of movement. *Curr. Opin. Neurobiol.* 8, 195–201
- 131 Cisek, P. and Kalaska, J.F. (2002) Simultaneous encoding of multiple potential reach directions in dorsal premotor cortex. *J. Neurophysiol.* 87, 1149–1154
- 132 Lebedev, M.A. and Wise, S.P. (2001) Tuning for the orientation of spatial attention in dorsal premotor cortex. *Eur. J. Neurosci.* 13, 1002–1008
- 133 Boussaoud, D. *et al.* (1993) Effects of gaze on apparent visual responses of frontal cortex neurons. *Exp. Brain Res.* 93, 423–434
- 134 Ivry, R.B. (1996) The representation of temporal information in perception and motor control. *Curr. Opin. Neurobiol.* 6, 851–857
- 135 Leon, M.I. and Shadlen, M.N. (2003) Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* 38, 317–327
- 136 Roux, S. *et al.* (2003) Context-related representation of timing processes in monkey motor cortex. *Eur. J. Neurosci.* 18, 1011–1016
- 137 Matell, M.S. *et al.* (2003) Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behav. Neurosci.* 117, 760–773
- 138 O'Doherty, J. *et al.* (2005) Ensemble representation of time: interhemispheric communication involved? In *2005 Abstract Viewer and Itinerary Planner*, Program No. 402.5, Society for Neuroscience Online (<http://sfn.scholarone.com/>)
- 139 Breuer, T. *et al.* (2005) First observation of tool use in wild gorillas. *PLoS Biol.* 3, e380
- 140 Head, H. and Holmes, G. (1911) Sensory disturbances from cerebral lesion. *Brain* 34, 102–254
- 141 Iriki, A. *et al.* (1996) Coding of modified body schema during tool use by macaque postcentral neurones. *NeuroReport* 7, 2325–2330
- 142 Maravita, A. *et al.* (2003) Multisensory integration and the body schema: close to hand and within reach. *Curr. Biol.* 13, R531–R539
- 143 Gurfinkel, V.S. *et al.* (1991) Body scheme concept and motor control. Body scheme in the postural automatism regulation. In *Intellectual Processes and Their Modelling*, pp 24–53, Nauka
- 144 Maruishi, M. *et al.* (2004) Brain activation during manipulation of the myoelectric prosthetic hand: a functional magnetic resonance imaging study. *NeuroImage* 21, 1604–1611
- 145 Roux, F.E. *et al.* (2003) Cortical areas involved in virtual movement of phantom limbs: comparison with normal subjects. *Neurosurgery* 53, 1342–1352
- 146 Zacksenhouse, M. *et al.* (2005) Trends in firing rate statistics mirroring changes in test performance during training with brain machine interfaces. In *2005 Abstract Viewer and Itinerary Planner*, Program No. 402.4, Society for Neuroscience Online (<http://sfn.scholarone.com/>)
- 147 Todorov, E. and Jordan, M.I. (2002) Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235
- 148 Scott, S.H. (2004) Optimal feedback control and the neural basis of volitional motor control. *Nat. Rev. Neurosci.* 5, 532–546
- 149 Harris, C.M. and Wolpert, D.M. (1998) Signal-dependent noise determines motor planning. *Nature* 394, 780–784
- 150 Cohen, M.R. and Newsome, W.T. (2004) What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.* 14, 169–177
- 151 Romo, R. *et al.* (2000) Sensing without touching: psychophysical performance based on cortical microstimulation. *Neuron* 26, 273–278
- 152 Sandler, A. *et al.* (2004) Primate somatosensorimotor learning: examining cue-related, association-related and motor-related responses in several cortical areas. In *2004 Abstract Viewer and Itinerary Planner*, Program No. 884.7, Society for Neuroscience Online (<http://sfn.scholarone.com/>)
- 153 Fitzsimmons, N.A. *et al.* (2005) Long-term behavioral improvements in a reaching task cued by microstimulation of the primary somatosensory cortex. In *2005 Abstract Viewer and Itinerary Planner*, Program No. 402.7, Society for Neuroscience Online (<http://sfn.scholarone.com/>)

Articles of interest in *Current Opinion* journals

The development and modulation of nociceptive circuitry

Xu Zhang and Lan Bao

Current Opinion in Neurobiology DOI: 10.1016/j.conb.2006.06.002

Seeing sounds: visual and auditory interactions in the brain

David A. Bulkin and Jennifer M. Groh

Current Opinion in Neurobiology DOI: 10.1016/j.conb.2006.06.008

The decoding of electrosensory systems

Eric S. Fortune

Current Opinion in Neurobiology DOI: 10.1016/j.conb.2006.06.006

Touching on somatosensory specializations in mammals

Kenneth C. Catania and Erin C. Henry

Current Opinion in Neurobiology DOI: 10.1016/j.conb.2006.06.010

Lifelong learning: ocular dominance plasticity in mouse visual cortex

Sonja B. Hofer, Thomas D. Mrsic-Flogel, Tobias Bonhoeffer and Mark Hübener

Current Opinion in Neurobiology DOI: 10.1016/j.conb.2006.06.007

Molecular Trojan horses for blood-brain barrier drug delivery

William M. Pardridge

Current Opinion in Pharmacology DOI: 10.1016/j.coph.2006.06.001

ARTICLES

Neuronal ensemble control of prosthetic devices by a human with tetraplegia

Leigh R. Hochberg^{1,2,4}, Mijail D. Serruya^{2,3}, Gerhard M. Friehs^{5,6}, Jon A. Mukand^{7,8}, Maryam Saleh^{9†}, Abraham H. Caplan⁹, Almut Branner¹⁰, David Chen¹¹, Richard D. Penn¹² & John P. Donoghue^{2,9}

Neuromotor prostheses (NMPs) aim to replace or restore lost motor functions in paralysed humans by routing movement-related signals from the brain, around damaged parts of the nervous system, to external effectors. To translate preclinical results from intact animals to a clinically useful NMP, movement signals must persist in cortex after spinal cord injury and be engaged by movement intent when sensory inputs and limb movement are long absent. Furthermore, NMPs would require that intention-driven neuronal activity be converted into a control signal that enables useful tasks. Here we show initial results for a tetraplegic human (MN) using a pilot NMP. Neuronal ensemble activity recorded through a 96-microelectrode array implanted in primary motor cortex demonstrated that intended hand motion modulates cortical spiking patterns three years after spinal cord injury. Decoders were created, providing a 'neural cursor' with which MN opened simulated e-mail and operated devices such as a television, even while conversing. Furthermore, MN used neural control to open and close a prosthetic hand, and perform rudimentary actions with a multi-jointed robotic arm. These early results suggest that NMPs based upon intracortical neuronal ensemble spiking activity could provide a valuable new neurotechnology to restore independence for humans with paralysis.

Hundreds of thousands of people suffer from forms of motor impairment in which intact movement-related areas of the brain cannot generate movements because of damage to the spinal cord, nerves, or muscles¹. Paralysing disorders profoundly limit independence, mobility and communication. Current assistive technologies rely on devices for which an extant function provides a signal that substitutes for missing actions. For example, cameras can monitor eye movements that can be used to point a computer cursor². Although these surrogate devices have been available for some time, they are typically limited in utility, cumbersome to maintain, and disruptive of natural actions. For instance, gaze towards objects of interest disrupts eye-based control. By contrast, an NMP is a type of brain-computer interface (BCI) that can guide movement by harnessing the existing neural substrate for that action—that is, neuronal activity patterns in motor areas. An ideal NMP would provide a safe, unobtrusive and reliable signal from the disconnected motor area that could restore lost function. Neurons in the primary motor cortex (MI) arm area of monkeys, for example, provide information about intended arm reaching trajectories^{3–5}, but this command signal would work for an NMP only if neural signals persist and could be engaged by intention in paralysed humans.

In concept, NMPs require a sensor to detect the activity of multiple neurons, a decoder to translate ensemble firing patterns into motor commands, and, typically, a computer gateway to engage effectors. BrainGate (Cyberkinetics, Inc.) is an NMP system under development

and in pilot trials in people with tetraparesis from spinal cord injury, brainstem stroke, muscular dystrophy, or amyotrophic lateral sclerosis. Currently, this system consists of a chronically implanted sensor and external signal processors developed from preclinical animal studies (see Methods)^{6–8}. The participant described in this report, the first in the BrainGate trial, is a 25-yr-old male (MN) who sustained a knife wound in 2001 that transected the spinal cord between cervical vertebrae C3–C4, resulting in complete tetraplegia (C4 ASIA A)⁹. The array was implanted in June 2004 into the MI arm area 'knob'¹⁰, as identified on pre-operative magnetic resonance imaging (MRI) (Fig. 1c). Post-operative recovery was uneventful. The data presented here are derived from 57 consecutive recording sessions from 14 July 2004 to 12 April 2005 (9 months).

Signal quality and variety

Action potentials were readily observable on multiple electrodes, indicating that MI neural spiking persists three years after SCI, as suggested indirectly by functional MRI data^{11–14}. Recorded signals ranged from qualitatively well-isolated single neurons to mixtures of a few different waveforms (Fig. 2a). Different waveform shapes were identified visually, using standard time-amplitude windows, but there was no further attempt to distinguish between well isolated and intermixed waveforms, both of which we refer to in this report as 'units'. An average of 26.9 ± 14.2 units were observed each day (range 3–57), with mean peak-to-peak spike amplitudes of $76.4 \pm 25.0 \mu\text{V}$ (mean $\pm$ s.d., $n = 56$ sessions) (see Supplementary

¹Department of Neurology, Massachusetts General Hospital, Brigham and Women's Hospital, and Spaulding Rehabilitation Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA. ²Department of Neuroscience and Brain Science Program, and ³Department of Engineering, Brown University, PO Box 1953, Providence, Rhode Island 02912, USA. ⁴Center for Restorative and Regenerative Medicine, Rehabilitation Research and Development Service, Department of Veterans Affairs, Veterans Health Administration, 830 Chalkstone Avenue, Providence, Rhode Island 02908, USA. ⁵Department of Clinical Neurosciences (Neurosurgery), Brown University, and ⁶Department of Neurosurgery, Rhode Island Hospital, 120 Dudley Street, Suite 103, Providence, Rhode Island 02905, USA. ⁷Department of Rehabilitation Medicine, Brown University, 593 Eddy Street, Providence, Rhode Island 02903, USA. ⁸Sargent Rehabilitation Center, 800 Quaker Lane, Warwick, Rhode Island 02818, USA. ⁹Cyberkinetics Neurotechnology Systems, Inc., 100 Foxborough Boulevard-Suite 240, Foxborough, Massachusetts 02035, USA. ¹⁰Cyberkinetics Neurotechnology Systems, Inc., 391 Chipeta Way, Suite G, Salt Lake City, Utah 84108, USA. ¹¹Department of Physical Medicine and Rehabilitation, Rehabilitation Institute of Chicago, 345 E. Superior Street, 1146, Chicago, Illinois 60611, USA. ¹²Department of Neurosurgery, University of Chicago Hospitals, 5841 S. Maryland Avenue, MC3026, Chicago, Illinois 60637, USA. †Present address: Graduate Program in Computational Neuroscience, University of Chicago, Chicago, Illinois 60637, USA.

Information)⁸. Although details of local field potentials (LFPs) will be described in a subsequent report, we note that LFPs, which could be recorded simultaneously with spikes, resembled those observed in intact monkeys (Fig. 2b)¹⁵. A notable decrease in the number of recorded units was seen approximately 6.5 months after implantation and thereafter. Upon approval of a clinical protocol change at 10 months, which permitted impedance measurements, we observed a low impedance in 54 of the electrodes, consistent with a physical short circuit to ground in the array, cable, or connector, but not consistent with a biological event (for example, gliosis). The precise reason for this physical change remains under investigation.

Since this report was written, we added a second trial participant to the study, a 55-yr-old man who has had complete spinal cord injury at C4 since 1999. Recordings were collected starting in the seventh month after implant, after making an electrical contacts repair in the pedestal connector. We recorded an average of 53.2 ± 6.3 units per session during trial months 7–10, again demonstrating the presence of neural activity lasting many months. Another technical issue causing abrupt signal loss at most electrodes, which may be related to the original repair, occurred at month 11 in participant 2; the reason for this change is being evaluated.

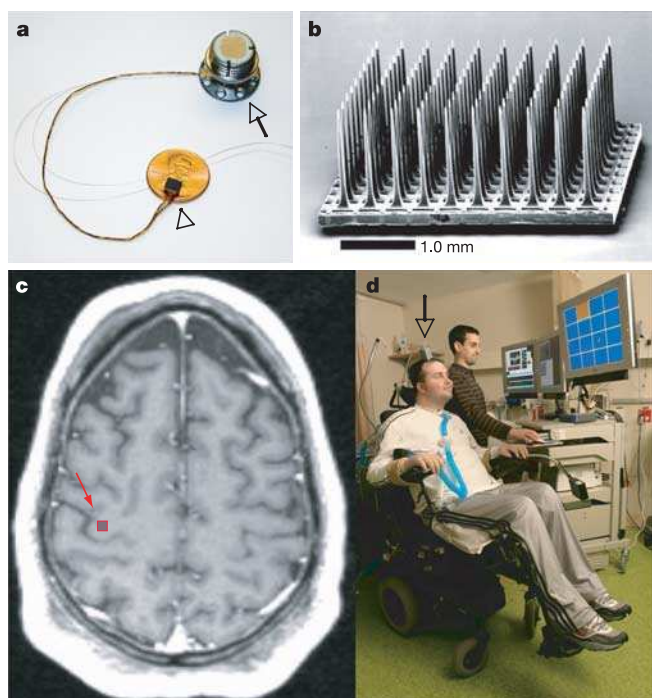


Figure 1 | Intracortical sensor and placement, participant 1. **a**, The BrainGate sensor (arrowhead), resting on a US penny, connected by a 13-cm ribbon cable to the percutaneous Ti pedestal (arrow), which is secured to the skull. Neural signals are recorded while the pedestal is connected to the remainder of the BrainGate system (seen in **d**). **b**, Scanning electron micrograph of the 100-electrode sensor, 96 of which are available for neural recording. Individual electrodes are 1-mm long and spaced 400 μm apart, in a 10×10 grid. **c**, Pre-operative axial T1-weighted MRI of the brain of participant 1. The arm/hand 'knob' of the right precentral gyrus (red arrow) corresponds to the approximate location of the sensor implant site. A scaled projection of the 4×4 -mm array onto the precentral knob is outlined in red. **d**, The first participant in the BrainGate trial (MN). He is sitting in a wheelchair, mechanically ventilated through a tracheostomy. The grey box (arrow) connected to the percutaneous pedestal contains amplifier and signal conditioning hardware; cabling brings the amplified neural signals to computers sitting beside the participant. He is looking at the monitor, directing the neural cursor towards the orange square in this 16-target 'grid' task. A technician appears (A.H.C.) behind the participant.

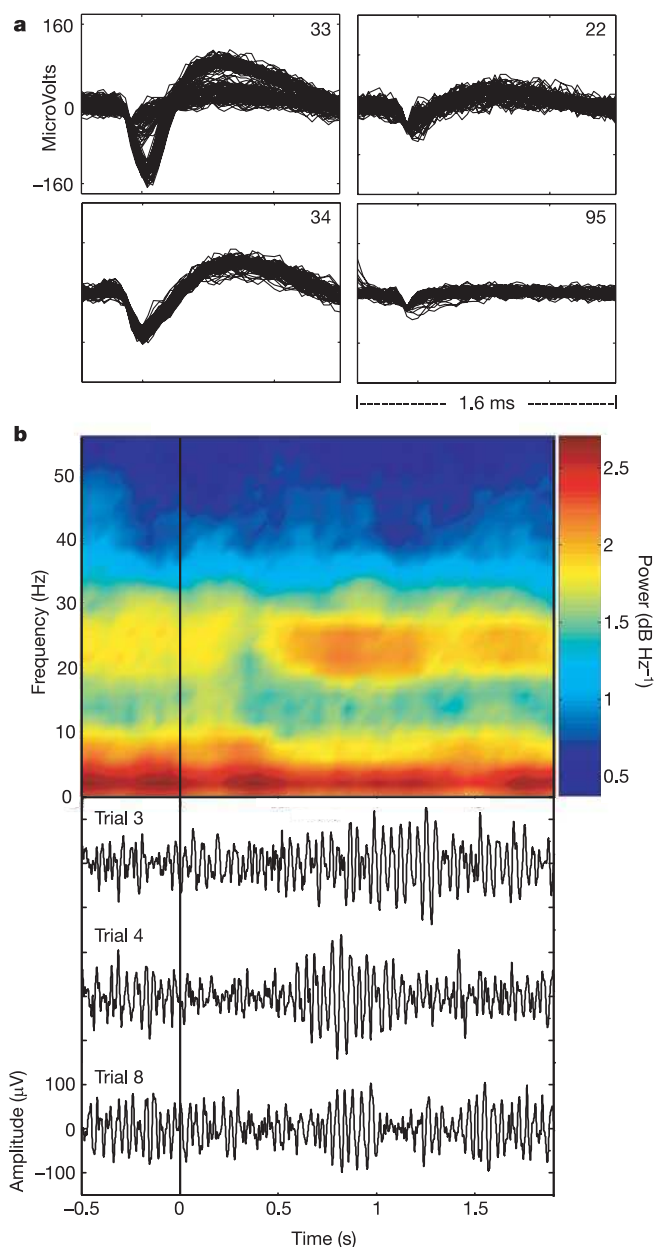


Figure 2 | Electrical recordings from a sample of four electrodes. **a**, Discriminated neural activity at electrodes 33, 34, 22, 95 ($n = 80$ superimposed action potentials for each unit). On electrode 33, two neuronal units could be reliably discriminated with peak to peak amplitudes of 206 and 56 μV , respectively. For electrode 34, a single unit is displayed. Electrode 22 illustrates a low-amplitude discriminated signal. Electrode 95 shows triggered noise. Data are from trial day 90 (90 days after array placement). **b**, Local field potentials during neural cursor control. In the bottom panel, three traces of electrical recording (bandpass: 10–100 Hz) from one electrode are shown 0.5 s before and 1.9 s after the go cue instructing MN to move the cursor from the centre position to a target at the top of the screen. In the top panel, a Thomson multi-taper time frequency analysis on each trial data segment was performed. This was done by sliding a 0.3-s window every 0.05 s, using a spectral resolution of 10 Hz. These power spectrograms were averaged across 20 trials to create the resulting pseudocolour power spectral density (PSD) plot. The diagram is aligned such that each point in the PSD plot corresponds to a time window 150 ms before and after an LFP. In the 20–30-Hz band, a decrease in power is seen approximately 300 ms after the go cue, followed by an increase in power from 550–1,200 ms after the go cue, which can also be appreciated in the raw, single trial data below.

Modulation by intent

Imagined limb motions modulated neural firing on multiple electrodes, upon request, beginning at the first experimental session. Modulation was evaluated during four consecutive sessions when MN was asked to imagine a series of movements. This series revealed a rich variety of firing modulations largely consistent with patterns observed in monkey MI¹⁶. Importantly, this activity was evoked by imagined actions in this participant with cervical spinal cord injury. Figure 3a illustrates how certain neurons are selective for one imagined action (hands together/apart), whereas others recorded simultaneously are engaged by different imagined actions (elbow or wrist). This diversity includes neurons that fired with imagined hand or distal arm actions (for example, hand open/close, Fig. 3c) and those that fired during shoulder movements that were actually performed (see also Supplementary Fig. 1). Non-selective neurons, active with the onset of any imagined upper extremity action (Fig. 3b), were also observed. As shown in Fig. 3c and Supplementary

Fig. 1, neurons fired in a relatively time-locked manner upon the request to imagine action. These results demonstrate a rich heterogeneity of firing patterns within a limited sample from a small MI region. This diversity is useful in creating a flexible control signal.

Linear filter construction

Use of MI neuronal ensemble activity as a control signal by persons with paralysis requires a novel approach to establishing a transform (filter function) between firing patterns and intended action. For each session, units were used to create a filter (see Methods) to provide a two-dimensional output signal displayed as cursor position on a monitor. A simple linear filter algorithm, identical to that used in intact monkeys, was used to create filters¹⁷. Unlike most studies performed using intact monkeys, where hand motions are known and kinematics are measured directly, we predicted MN's intended hand movements on the basis of

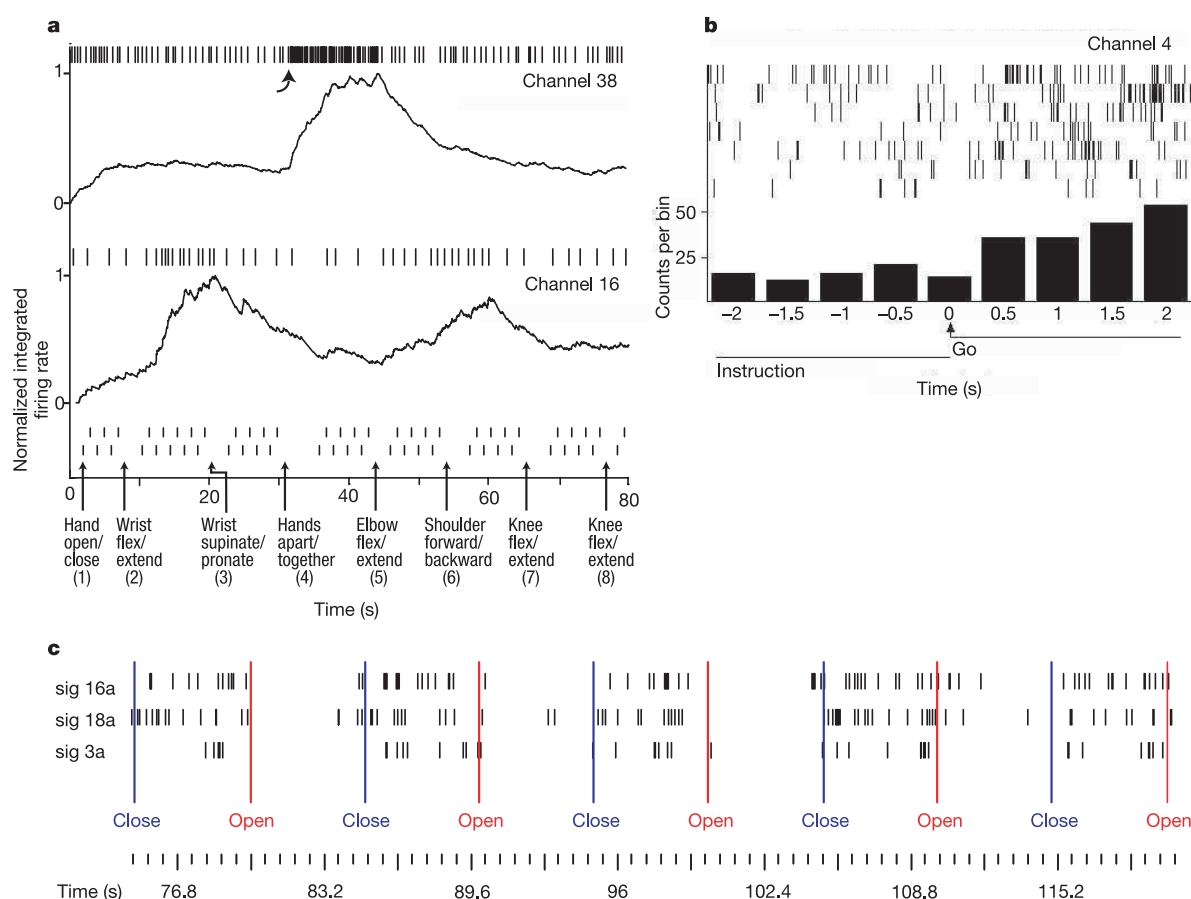


Figure 3 | Neuronal selectivity for imagined and performed movements.

a, Over an 80-s period, MN was asked to imagine performing a series of left limb movements (which are described on the abscissa). Movement instruction time is indicated by a vertical arrow; the go cue for each alternating movement, presented as text on the video monitor, is indicated by a small vertical hash mark. Spiking activity of two simultaneously recorded units is displayed. Rasters indicate the time of each spike (thinned for visual clarity; every third spike is shown). Normalized, integrated firing rates (R) appear beneath each raster, as derived by the equation $R = [(R_{-1} + n)(1 - e^{-b/\tau})]$, where R_{-1} is the previous bin's integrated firing rate value, n = the number of spikes in the current bin, b = bin width, and τ = time constant; bin width = 50-ms window, time constant = 10 ms (adapted from ref. 28); normalization is achieved by dividing by the maximum integrated firing rate from each unit's spike train over the time period displayed. The top unit (channel 38) increases its firing rate (curved arrow) with the instruction to move both hands apart/together. The bottom, simultaneously recorded unit (channel 16) is activated most clearly after the

instruction to flex/extend the wrist and to flex/extend the elbow or move the shoulder anteriorly and posteriorly. All movements are imagined except for shoulder movement, which MN actually performed. **b**, Go-cue-related activity modulation for a neuron recorded simultaneously with those in **a**. Each raster line is centred about the go cue, which requests that the patient imagine a movement; the seven raster lines represent the epochs surrounding each of the seven different movements in sequence **a**. The histogram displays the total number of spikes seen in each 500-ms bin. This neuron increased its firing rate during most imagined movement epochs, but demonstrated poor instruction selectivity compared to the neurons presented in **a**. Data are from day 161. **c**, Hand-instruction-related modulation for three simultaneously recorded neurons. MN was cued to open and close his hand by text instruction, presented on the screen. Go cues are indicated. Each vertical tick represents one action potential (spike). An increase in these neurons' firing rates is directly indicative of the intention to close the hand. Data are from day 90.

instructed actions (instruction-based algorithms have also been reported in one study of intact monkeys¹⁸). Thus, for filter building, MN was asked to imagine manually tracking a 'technician's cursor' that was actually being moved by a technician-operated mouse through a succession of randomly placed visual targets (see Methods). The filter function was used to decode activity and drive a 'neural cursor'.

MI activity during neural cursor control

Features of neurons during neural cursor control resembled those expected from MI. Neurons in MI of intact monkeys characteristically begin to modulate their firing before movement onset and activity is tuned to hand movement direction^{19–21}. To compare this neural activity with MI of a human with spinal cord injury, MN performed a step-tracking, 'centre-out' task using the neural cursor. The task requires that the neural cursor be moved from a centre target to one of four radially displaced targets (screen location: up, down, left, right; see Supplementary Video 1). For each of six sessions, MN performed this task by imagining hand motion (see Methods) as soon as the target cue appeared. The task was performed immediately after filter building without intervening practice. Timing and directional tuning features of MI neurons during imagined actions were consistent with those observed in MI of intact non-human primates. Figure 4 shows that spike-rate modulation occurs soon after the 'go' cue and that modulation varied by target location, as would be predicted for MI if actual arm motions were performed¹⁷. Furthermore, 66 out of 73 discriminated units (90.4%) significantly changed their firing rate in relation to the appearance of the go cue (Kolmogorov–Smirnov test, $\alpha = 0.05$, rate calculated over a sliding 1-s window, overlapping every 0.05 s; 60-s data set for each condition, $n = 3$ sessions). These results indicate that, even years after spinal cord injury and in the absence of kinaesthetic feedback and limb movement, MI neurons can still be actively engaged and encode

task-related information during the intention to move the limb ordinarily controlled by that MI region.

Quality of neural cursor control

Neural cursor position was significantly correlated with technician cursor position during the last block of the pursuit filter building task (x coordinate $r^2 = 0.56 \pm 0.18$ and y coordinate $r^2 = 0.45 \pm 0.15$, $n = 6$ sessions, Fig. 5). These correlations are similar or better than those seen in intact monkeys when linear filters were used to predict real-time hand position from MI neuronal ensembles^{5,22}. The neural cursor could be directed towards targets with a form qualitatively similar to that seen for intact monkeys using closed-loop neural control^{17,18,23}. As in intact monkeys, neural cursor motion had underlying instabilities and variable oscillatory components compared to hand motions of able-bodied individuals. Continuous neural cursor motion with the linear filter made cursor fixation at a single location difficult to achieve.

Data from the centre-out task were used to evaluate the speed and accuracy of cursor control, which are essential design parameters for any future practical NMP. As shown in Fig. 6, the participant correctly acquired 73–95% of targets (control 6.5%; $n = 80$, paired t -test, $P < 0.0001$, see Methods) when measured in a series of six sessions (see also Supplementary Fig. 2). Performance errors reflected both instabilities in cursor direction control and the ability to hold at the target location. Mean time to target was 2.51 ± 0.16 s ($\pm$ s.e.m.) for successfully acquired targets. Although the best 13% of MN's trials were within the range consistently achieved by able-bodied controls using a computer mouse ($n = 3$, mean 1.06 ± 0.08 s ($\pm$ s.e.m.)), the distribution of times for MN using neural control is skewed to longer acquisition times (Fig. 6b). Effective use of the neural cursor in more complex spatial control tasks was evident when MN directed the cursor to randomly placed targets while attempting to avoid obstacles in the cursor's path (see Supplementary Video 5).

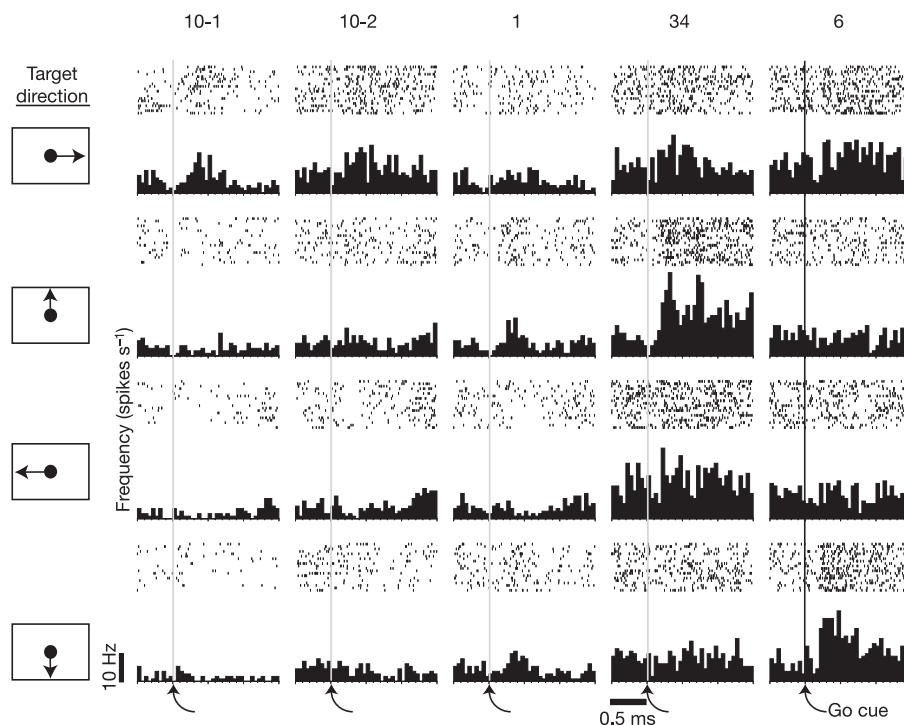


Figure 4 | Directional tuning during centre-out task. Peristimulus time histograms show spike rates for five neurons recorded simultaneously during the performance of a four-direction centre-out task (day 90) in which MN used the neural cursor to acquire a target presented at the right, top, left, or bottom of the screen. Twenty trials are displayed for each target location. Increases in activity after the go cue demonstrate movement-intention-

related modulation. Each column shows the firing of one unit in the four directions, aligned on the cue to move. Note, for example, the time-locked increase in firing of unit 6 when MN was cued to move the cursor downward (lower right corner) and the lack of change in firing rate for upward instruction. Changes in firing across the five neurons reveal directional tuning.

Further information concerning spatial and temporal accuracy was obtained from a 'grid task' (see Supplementary Information). As shown in Supplementary Video 8, participant 2 also performed the centre-out task, highlighting the ability for this second participant with spinal cord injury to modulate his motor cortical activity voluntarily for external device control. This participant's neural control, however, was generally less accurate and consistent than MN's; we are investigating the extent to which technical or other factors may underlie this performance. This and other participants' data will be reported in subsequent manuscripts.

Although he is tetraplegic, MN retains shoulder, neck and head mobility, and some recorded MI cells fired during shoulder movement

(see Supplementary Fig. 1). NMPs will nearly always operate in the context of some existing movement capabilities (given the considerable variability of remaining sensory and motor functions in people with CNS injury); thus, during filter building and use MN was not asked to remain completely still, and he sometimes moved his head or neck. It is thus possible that retained movements influenced cursor action, much as such movements are known to influence activity correlated with ordinary hand actions²⁴. These still-intact movements, however, did not appear to be essential for MN's cursor control, as can be appreciated particularly when he played 'Neural Pong' (Supplementary Video 4). In this video, movement of the head or the shoulders sometimes accompanied cursor control, but at other times MN moved the cursor purposefully while his head remained stationary, and during other epochs his head moved but the cursor motion appeared to be unrelated to head movement. We compared neural cursor position in Supplementary Video 4 to head position (both assessed by coordinates on a video playback monitor) and found no consistent relationship ($r^2 = 0.063$). This finding is inconsistent with a unique causal relation between head and cursor control. In addition, it was our repeated observation that, at least some of the time, MN performed neural control (and open-loop) tasks without moving his shoulder. Thus, based upon our instructions and MN's self-reporting that he was actively imagining arm action, we conclude that cursor motion is under the guidance of imagined/intended actions.

Direct control of prosthetic devices

Continuous computer cursor control could be used to provide many valuable new outputs for a person with paralysis to carry out

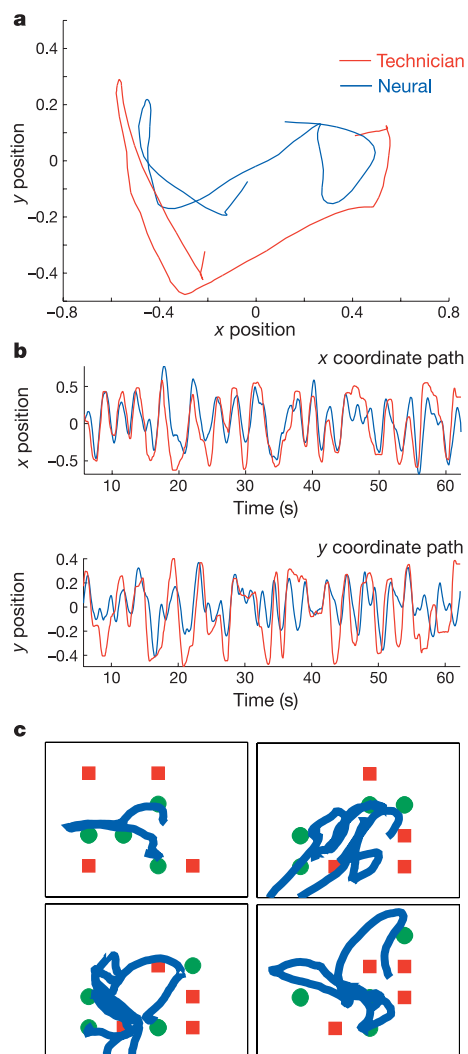


Figure 5 | Reconstruction of neural cursor position during pursuit tracking. **a**, Technician (red line) and neural cursor (blue line) paths during a 5-s epoch during which MN was asked to track the technician cursor with his neural cursor in real time. MN was able to track the general direction of the technician cursor with the neural cursor, changing directions quickly, while having some difficulty in overlaying the cursors precisely. Trial day 90. **b**, x, y position control over time during one tracking trial (last 1-min epoch of filter building). The top panel displays the x coordinate position of the technician cursor (red) and the neural cursor (blue); y positions for the same movement are shown in the bottom panel. **c**, Neural cursor position during a target acquisition/obstacle avoidance task. The four panels represent the four epochs shown in Supplementary Video 5. Green circles indicate targets; red squares indicate obstacles. The thick blue line indicates the path taken by the neural cursor and illustrates the ability to avoid most obstacles and acquire most targets within a randomly arranged field. Data are from trial day 90.

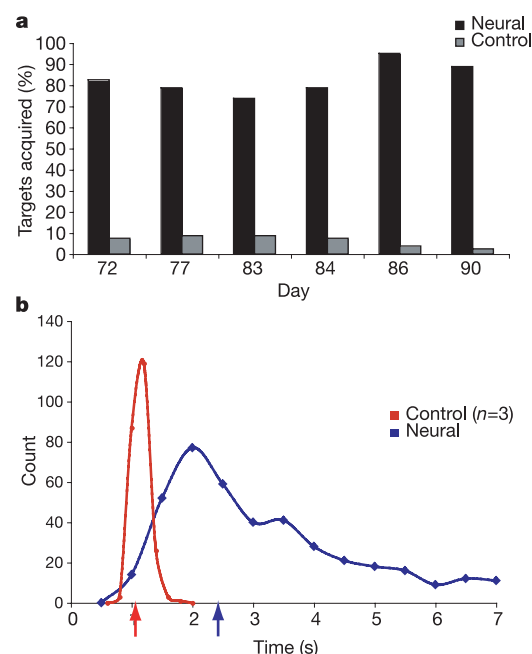


Figure 6 | Centre-out task performance. **a**, Target acquisition accuracy during the centre-out task. For each of six sessions, MN acquired between 73–95% of the radially placed targets. Control targets were not present on the monitor during task performance, but were marked as acquired if, during post-hoc analysis of the cursor movement, the cursor had traversed the location of one of the other three pseudo-randomly selected targets before the correct target (see Supplementary Video 1). Data from days 72, 77, 83, 84, 86, 90 are shown. **b**, Time-to-target performance during centre-out task for MN (blue) and three able-bodied controls (red). Only successful target acquisitions in <7 s are shown for MN. Arrows on the abscissa represent median times to target for each distribution. Controls' performances ($n = 3$ controls, 80 trials each) are collapsed into 0.2-s bins. MN's performance (398 trials) is collapsed into 0.5-s bins for visual clarity.

activities of daily living. Such a control signal could be used not only to direct computer software, but also to operate external physical assistive devices. MN tested his ability to perform potentially useful actions in a series of demonstrations.

MN used a simplified computer interface to open simulated e-mail and to draw an approximately circular figure using a paint program (Supplementary Video 2). Using the neural cursor coupled to a simple hardware interface, he adjusted the volume, channel and power to his television. He was also able to play video games such as Neural Pong (see Supplementary Videos 3–5). Control of two robotic devices was achieved, allowing MN to manipulate directly the environment. In one example, neural output was coupled to a prosthetic hand (Liberating Technologies, Inc.; see Methods) and MN was able to open and close the hand under volitional control (see Supplementary Video 6). Although only a one-dimensional form of proportional control, he achieved this action after a few trials while looking at the hand and without requiring feedback from the cursor display. Lastly, MN used a simple multi-jointed robotic limb to grasp an object and transport it from one location to another (see Supplementary Video 7 and Supplementary Information). These demonstrations illustrate transfer of control directly to a physical device without depending on continued viewing of computer cursor feedback, and suggest that manipulation of the environment to enable activities such as self-paced eating, as well as other goal-directed actions, could be achieved in tetraplegic humans.

Notably, each of these tasks was achieved rapidly and could be performed while the participant was conversing. Thus, the MI-based NMP may have the property of allowing external device control with little more disruption than encountered in able-bodied humans when they are using their arms or hands and simultaneously carrying out other motor or cognitive functions.

Discussion

The research shown here from the first participant in an ongoing pilot clinical trial provides initial evidence that a human unable to move or sense his limbs can operate a NMP using MI neuronal ensemble spiking activity as a control source. In addition, we demonstrate that neural spiking remains in the MI arm area and can be modulated by intention years after spinal cord injury. These findings provide a number of novel insights concerning cortical function and the impact of spinal cord injury in humans (see Supplementary Information).

Although MN used direct neural control to perform reasonably challenging tasks, the level of the control is considerably less than that of an able-bodied person using a manually controlled computer mouse. A number of factors might affect control, including: (1) the small set of randomly selected neurons recorded by this single array, compared to the very large number usually engaged; (2) the influence of spinal cord injury mechanisms or duration since injury; (3) the cortical layers recorded; (4) our approach to filter building; (5) the nature of the linear filter (as compared, for example, to Kalman filters²², adaptive algorithms or support vector machine approaches²⁵, not yet tested in our work); (6) attention and motivational state during filter building or control; and (7) the user interface. Changes in the recorded population across days may also contribute to both variability and instability of control. Shifting ensembles may result from small motions of the array or through other poorly understood mechanisms. Despite these variables, it is important to note that useful filters could be created daily from that neural population, and that advances in knowledge and technology are likely to improve recording and decoding. For example, cursor control might be enhanced further by adaptive algorithms or more selective choice of neural signals. Combinations of spiking activity and simultaneously recorded LFPs would provide additional signals that might improve performance.

Efforts to optimize performance may help to identify the design

criteria for a clinically useful interface for paralysed humans. For example, electroencephalogram (EEG)-based control systems have been improved by limiting trial time, re-centring the cursor location after each trial, and stopping cursor motion as soon as targets were hit²⁶. By contrast, our participant had to maintain cursor control at all times without these interface enhancements. However, we found that trial success was greater when target dwell time was decreased (grid task). Incorporation of additional feedback (for example, visual, auditory, somatosensory) may also be useful in enhancing performance²⁷.

Cursor and external device control may also be improved through learning²⁸. MI of intact animals shows learning-related plasticity^{29–32}, and such plasticity undoubtedly contributes to the acquisition of direct cortical neuronal control of an NMP. Thus, we expect that learning will further improve control, as has been reported in monkeys¹⁸. The circle drawing task (Supplementary Video 2), for example, provides some evidence of improved performance with brief practice, but a better understanding of learning will require additional, more systematic investigation.

Other human BCIs are under development³³. Foremost for comparison here is one by Kennedy and collaborators, who have reported results with their neurotrophic (cone) electrode for three humans: one with amyotrophic lateral sclerosis³⁴, one with brainstem stroke^{35,36}, and one with mitochondrial myopathy³⁶ (see Supplementary Information). In contrast to previous studies, here we present evidence that cortical ensemble patterns can be decoded and used for real-time, two-dimensional control of a computer cursor and other external devices by a person with spinal cord injury. Scalp-based EEG-driven BCIs also have been tested with some success in people with amyotrophic lateral sclerosis, spinal cord injury and other forms of paralysis^{26,27,37–39}. Such indirect systems use a substitute for motor ensembles to achieve control. Although two-dimensional control has been achieved in paraplegic patients using independent modulation of scalp-recorded signals²⁶, this system requires significant learning, concentration to the exclusion of other actions, and daily scalp sensor application, but notably, not surgical placement of the sensor. It also has limited scalability to multiple functions because two-dimensional tasks appear to engage all controllable signals. In contrast, the direct NMP can be used during natural activities such as speech, and requires minimal learning beyond filter creation. It is also plausible that NMPs could be scaled so that parallel commands could be derived simultaneously from multiple sensors each in separate cortical regions. If achieved, relatively independent outputs potentially could emanate bilaterally from arm and leg areas to allow for the reanimation of paralysed limbs via functional electrical stimulation devices.

The relative merits of EEG, transcranial⁴⁰, electrocorticographic (ECoG)^{41–43} and intracortically based BCIs (NMPs) continue to be explored and expanded^{32,33}. Myriad issues including efficacy, safety (particularly with regards to surgery), reliability, longevity, required training and support, cosmesis, cost and availability will undoubtedly affect BCI success; individuals with impairments in communication and/or mobility may prioritize these factors differently when choosing a BCI, and it is possible that combinations of techniques will provide most useful restoration of control. Current BCIs (including the NMP reported here) depend upon the insights and assistance of trained experts. The need for this assistance must be eliminated through system automation.

The goal of NMPs and other BCI research is to create safe, reliable and unobtrusive neural interfaces to devices that will restore the communication, mobility, and independence of paralysed humans. NMPs could provide significant advances over current technologies because they engage natural substrates for control (that is, the MI arm/hand area), do not encumber other actions, and do not require extensive learning. It will be important to collect results from additional participants in order to understand the generality of these statements. Improvements in the decoder and user interface

are likely to provide control closer to that achieved by able-bodied humans. The potential to use implanted sensors to 're-route' neural signals may also prove valuable in stroke and other neurological disorders where pathways are disconnected. The present NMP system incorporates a transcutaneous connection that tethers a participant to a bulky cart and requires operation by a trained technician. A wireless, implantable and miniaturized system combined with automation will be required for practical use. Emerging and available technologies appear to be sufficient to overcome these obstacles, although the challenges of creating a fully implantable system may be formidable. The current data provide initial proof of concept that spiking activity from neuronal ensembles can provide a control signal after spinal cord injury sufficient to perform at least basic operations for a human with tetraplegia, justifying further engineering efforts towards a human NMP.

METHODS

See Supplementary Information for additional Methods.

Approval for these studies was granted by the US Food and Drug Administration (Investigational Device Exemption) and the Rhode Island Hospital, New England, Rehabilitation Institute of Chicago, University of Chicago, and Spaulding Rehabilitation Hospital Institutional Review Boards. The participant described in this report has provided permission for photographs, videos and portions of his protected health information to be published for scientific and educational purposes. After completion of informed consent and medical and surgical screening procedures, the 4 × 4-mm array of electrodes was implanted into the motor cortex using a pneumatic insertion technique^{8,44}. Details of the human surgical procedure are in preparation for publication.

BrainGate system. The sensor is a 10 × 10 array of silicon microelectrodes that protrude 1 mm from a 4 × 4-mm platform (Fig. 1b). At manufacture, electrodes had an impedance of $322 \pm 138 \text{ k}\Omega$ (mean $\pm$ s.d.). The array was implanted onto the surface of the MI arm/hand region; electrodes penetrate into the cortex to attempt to record neurons in intermediate layers. Recorded electrical signals pass externally through a Ti percutaneous connector, which is secured to the skull. Cabling attached to the connector during recording sessions routes signals to external amplifiers and then to a series of computers in a cart that process the signals and convert them into an output (the neural cursor) that can be viewed by the participant on a computer monitor (Fig. 1d). Currently, the system must be set up and managed by an experienced technician.

Recording sessions. Research sessions were scheduled at least once per week at the participant's residence. Sessions could be cancelled (for example, due to participant request or schedule conflict) or ended early at the participant's request. Sessions would commence with neural recording and spike discrimination, followed by filter building and structured clinical end-point (cursor control) trials. Performance of video games and external device control demonstrations followed. The electrodes and neural signals selected immediately before filter building remained constant for any given session's cursor control trials; additional neural signals (detected either with alternative spike discrimination methods, or as may have become evident during the cursor control trials) could be used during subsequent video game or external device use.

Spike discrimination. Units were manually discriminated by a technician using visual features to place time-amplitude windows on waveforms displayed within 1.6-ms windows triggered when the signal crossed a manually adjusted threshold (Fig. 2) (Cyberkinetics Central Software) while the participant was at rest. We applied no objective criteria to determine whether any particular unit (or admixture of units) recorded on one day was the same as that recorded on a subsequent day. Smaller amplitude signals were likely to have consisted of admixtures of spikes from >1 neuron. These signals were nevertheless used for sessions.

Filter building. For each session, single and multiunit data were used to create a filter to generate a two-dimensional output signal. During filter building, MN was asked to imagine moving his hand as if he were controlling a computer mouse. Specifically, he was asked to imagine tracking a cursor on the computer screen; this technician cursor was moved (by the technician) through a succession of randomly positioned targets. During the first 4 min of imagined tracking, only the technician cursor and targets were visible on the screen. An initial filter was built with the neural activity collected during this epoch, then allowing a neural cursor to be placed on the screen. MN continued to track the technician cursor with his neural cursor over a series of four 1-min blocks. The filter was updated at the end of each block. To create the final filter for the research session, neural data were used from only the four neural cursor blocks.

Linear filters were constructed from a response matrix containing the firing

rate over a 1-s history for each neuron (twenty 50-ms bins), and regressing this matrix onto technician-cursor position using a pseudoinverse technique. The least-squares formulation comprises a closed-form solution:

$$\mathbf{u} = \mathbf{R}\mathbf{f} = \mathbf{R}(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{k}$$

where $\mathbf{R}$ is the neural response matrix, $\mathbf{f}$ is the linear filter, $\mathbf{k}$ is the kinematic value (x, y position), $\mathbf{T}$ indicates the transpose of the matrix, and $\mathbf{u}$ is the reconstruction⁵. Software for displays was custom generated for this purpose and run on standard PC computers (Orbit Micro, dual Intel Pentium 4 Xeon processors, 2 GHz).

External devices. The robot arm was obtained from Lynxmotion, Lynx 5 Series. The electric prosthetic hand was generously provided by Liberating Technologies.

Received 22 March; accepted 6 June 2006.

1. Jackson, A. B., Dijkers, M., Devivo, M. J. & Poczatek, R. B. A demographic profile of new traumatic spinal cord injuries: change and stability over 30 years. *Arch. Phys. Med. Rehabil.* **85**, 1740–1748 (2004).
2. Thoumie, P. et al. Clinical and functional evaluation of a gaze controlled system for the severely handicapped. *Spinal Cord* **36**, 104–109 (1998).
3. Humphrey, D. R., Schmidt, E. M. & Thompson, W. D. Predicting measures of motor performance from multiple cortical spike trains. *Science* **170**, 758–762 (1970).
4. Schwartz, A. B., Taylor, D. M. & Tillery, S. I. Extraction algorithms for cortical control of arm prosthetics. *Curr. Opin. Neurobiol.* **11**, 701–707 (2001).
5. Paninski, L., Fellows, M. R., Hatsopoulos, N. G. & Donoghue, J. P. Spatiotemporal tuning of motor cortical neurons for hand position and velocity. *J. Neurophysiol.* **91**, 515–532 (2004).
6. Maynard, E. M., Nordhausen, C. T. & Normann, R. A. The Utah intracortical electrode array: a recording structure for potential brain-computer interfaces. *Electroencephalogr. Clin. Neurophysiol.* **102**, 228–239 (1997).
7. Guillory, K. S. & Normann, R. A. A 100-channel system for real time detection and storage of extracellular spike waveforms. *J. Neurosci. Methods* **91**, 21–29 (1999).
8. Suner, S., Fellows, M. R., Vargas-Irwin, C., Nakata, K. & Donoghue, J. P. Reliability of signals from chronically implanted, silicon-based electrode array in non-human primate primary motor cortex. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13**, 524–541 (2005).
9. Maynard, F. M. Jr et al. International standards for neurological and functional classification of spinal cord injury. American Spinal Injury Association. *Spinal Cord* **35**, 266–274 (1997).
10. Youssry, T. A. et al. Localization of the motor hand area to a knob on the precentral gyrus. A new landmark. *Brain* **120**, 141–157 (1997).
11. Lotze, M., Laubis-Herrmann, U., Topka, H., Erb, M. & Grodd, W. Reorganization in the primary motor cortex after spinal cord injury—A functional magnetic resonance (fMRI) study. *Restor. Neurol. Neurosci.* **14**, 183–187 (1999).
12. Shoham, S., Halgren, E., Maynard, E. M. & Normann, R. A. Motor-cortical activity in tetraplegics. *Nature* **413**, 793 (2001).
13. Sabbah, P. et al. Sensorimotor cortical activity in patients with complete spinal cord injury: a functional magnetic resonance imaging study. *J. Neurotrauma* **19**, 53–60 (2002).
14. Mikulic, D. J. et al. Adaptation in the motor cortex following cervical spinal cord injury. *Neurology* **58**, 794–801 (2002).
15. Donoghue, J. P., Sanes, J. N., Hatsopoulos, N. G. & Gaal, G. Neural discharge and local field potential oscillations in primate motor cortex during voluntary movements. *J. Neurophysiol.* **79**, 159–173 (1998).
16. Sanes, J. N. & Donoghue, J. P. Plasticity and primary motor cortex. *Annu. Rev. Neurosci.* **23**, 393–415 (2000).
17. Serruya, M. D., Hatsopoulos, N. G., Paninski, L., Fellows, M. R. & Donoghue, J. P. Instant neural control of a movement signal. *Nature* **416**, 141–142 (2002).
18. Taylor, D. M., Tillery, S. I. & Schwartz, A. B. Direct cortical control of 3D neuroprosthetic devices. *Science* **296**, 1829–1832 (2002).
19. Evarts, E. V. in *Handbook of Physiology* (ed. Brooks, V.) 1083–1120 (Williams and Wilkins, Baltimore, 1981).
20. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
21. Humphrey, D. R. & Tanji, J. in *Motor Control: Concepts and Issues* (eds Humphrey, D. R. & Freund, H. J.) 413–443 (John Wiley, London, 1991).
22. Wu, W. et al. Modeling and decoding motor cortical activity using a switching Kalman filter. *IEEE Trans. Biomed. Eng.* **51**, 933–942 (2004).
23. Carmena, J. M. et al. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol.* **1**, E42 (2003).
24. Baker, J. T., Donoghue, J. P. & Sanes, J. N. Gaze direction modulates finger movement activation patterns in human cerebral cortex. *J. Neurosci.* **19**, 10044–10052 (1999).
25. Olson, B. P., Si, J., Hu, J. & He, J. Closed-loop cortical control of direction using support vector machines. *IEEE Trans. Neural Syst. Rehabil. Eng.* **13**, 72–80 (2005).
26. Wolpaw, J. R. & McFarland, D. J. Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl Acad. Sci. USA* **101**, 17849–17854 (2004).

27. Birbaumer, N. et al. A spelling device for the paralysed. *Nature* **398**, 297–298 (1999).
28. Fetz, E. E. & Baker, M. A. Operantly conditioned patterns on precentral unit activity and correlated responses in adjacent cells and contralateral muscles. *J. Neurophysiol.* **36**, 179–204 (1973).
29. Fetz, E. E. & Finocchio, D. V. Operant conditioning of specific patterns of neural and muscular activity. *Science* **174**, 431–435 (1971).
30. Jacobs, K. M. & Donoghue, J. P. Reshaping the cortical motor map by unmasking latent intracortical connections. *Science* **251**, 944–947 (1991).
31. Gandolfo, F., Li, C., Benda, B. J., Schioppa, C. P. & Bizzi, E. Cortical correlates of learning in monkeys adapting to a new dynamical environment. *Proc. Natl Acad. Sci. USA* **97**, 2259–2263 (2000).
32. Schwartz, A. B. Cortical neural prosthetics. *Annu. Rev. Neurosci.* **27**, 487–507 (2004).
33. Hochberg, L. R. & Donoghue, J. P. Sensors for brain-computer interfaces. *IEEE Eng. Med. Biol.* (in the press).
34. Kennedy, P. R. & Bakay, R. A. Restoration of neural output from a paralyzed patient by a direct brain connection. *Neuroreport* **9**, 1707–1711 (1998).
35. Kennedy, P. R., Bakay, R. A., Moore, M. M., Adams, K. & Goldwithe, J. Direct control of a computer from the human central nervous system. *IEEE Trans. Rehabil. Eng.* **8**, 198–202 (2000).
36. Kennedy, P. R., Kirby, M. T., Moore, M. M., King, B. & Mallory, A. Computer control using human intracortical local field potentials. *IEEE Trans. Neural Syst. Rehabil. Eng.* **12**, 339–344 (2004).
37. Mason, S. G., Bohringer, R., Borisoff, J. F. & Birch, G. E. Real-time control of a video game with a direct brain–computer interface. *J. Clin. Neurophysiol.* **21**, 404–408 (2004).
38. Vaughan, T. M. et al. Brain-computer interface technology: a review of the Second International Meeting. *IEEE Trans. Neural Syst. Rehabil. Eng.* **11**, 94–109 (2003).
39. Muller-Putz, G. R., Scherer, R., Pfurtscheller, G. & Rupp, R. EEG-based neuroprosthesis control: a step towards clinical practice. *Neurosci. Lett.* **382**, 169–174 (2005).
40. Kennedy, P. et al. Using human extra-cortical local field potentials to control a switch. *J. Neural Eng.* **1**, 72–77 (2004).
41. Leuthardt, E. C., Schalk, G., Wolpaw, J. W., Ojemann, J. G. & Moran, D. W. A brain-computer interface using electrocorticographic signals in humans. *J. Neural Eng.* **1**, 63–71 (2004).
42. Graimann, B., Huggins, J. E., Levine, S. P. & Pfurtscheller, G. Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. *IEEE Trans. Biomed. Eng.* **51**, 954–962 (2004).
43. Lal, T. N., et al. in *Advances in Neural Information Processing Systems* (eds Saul, L. K., Weiss, Y. & Bottou, L.) (MIT Press, Cambridge, Massachusetts, 2005).
44. Rousche, P. J. & Normann, R. A. A method for pneumatically inserting an array of penetrating electrodes into cortical tissue. *Ann. Biomed. Eng.* **20**, 413–422 (1992).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank J. Joseph and D. Morris for assistance; L. Mermel for clinical planning advice; V. Zerris and M. Park for surgical assistance; G. Polykoff for clinical trial assistance; W. Truccolo for power spectral density analysis development; and the employees of Cyberkinetics for device engineering, manufacturing and clinical trial design and management. The authors also thank MN for his participation in this trial, and the nursing staff at his assisted care facility for their assistance. The authors are grateful to M. Serra and Sargent Rehabilitation Center, the study site, for administrative support. The photograph of MN (Fig. 1) is copyright 2005 Rick Friedman. This work was supported by Cyberkinetics Neurotechnology Systems, Inc.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to J.P.D. (john_donoghue@brown.edu).

The role of attention in visual processing

John H. R. Maunsell* and Erik P. Cook

Howard Hughes Medical Institute and Division of Neuroscience, Baylor College of Medicine, One Baylor Plaza, S-603, Houston, TX 77030, USA

Attention to a visual stimulus typically increases the responses of cortical neurons to that stimulus. Because many studies have shown a close relationship between the performance of individual neurons and behavioural performance of animal subjects, it is important to consider how attention affects this relationship. Measurements of behavioural and neuronal performance taken from rhesus monkeys while they performed a motion detection task with two attentional states show that attention alters the relationship between behaviour and neuronal response. Notably, attention affects the relationship differently in different cortical visual areas. This indicates that a close relationship between neuronal and behavioural performance on a given task persists over changes in attentional state only within limited regions of visual cortex.

Keywords: attention; visual cortex; single unit; macaque monkey; performance

1. INTRODUCTION

Attention is an important factor in sensory processing. Directing attention to a particular location in the visual field improves detection and discrimination, and shortens reaction times in that location relative to others (see Pashler 1998). When behavioural performance is near threshold, attention can reliably make the difference between success and failure. Attention also affects the responses of sensory neurons. Neurons typically respond more strongly when the stimulus that drives them is the focus of attention (see Braun *et al.* 2001). Behavioural performance depends on the activity of sensory neurons. Little is known, however, about how the behavioural effects of attention are related to the changes that it produces in sensory representations.

This question can be approached experimentally because there appears to be a close relationship between behavioural performance on a particular task and the activity of individual sensory neurons. When the sensitivity of individual neurons is compared with the behavioural capabilities of the organism, the ability of a neuron to report about a stimulus can be about as good as that of the observer, provided the task involves stimuli that are closely matched to the preferences of the neuron in question (reviewed by Parker & Newsome 1998). This close correspondence between neuronal and behavioural performance indicates that behavioural performance depends on, and closely follows, the selectivity of sensory neurons (Barlow 1985).

Given this correspondence between neuronal and behavioural performance, and the observation that attention changes the responses of sensory neurons, it is natural to ask whether effects of attention on behavioural perform-

ance might be fully explained by changes that attention causes in the responses of sensory neurons. To address this issue, it is necessary to consider the magnitude and quality of effects that attention has on neuronal responses. In the sections that follow, we consider attentional effects on neuronal responses in the context of visual cortex, for which the most detailed and extensive data are available. We then describe recent experiments from our laboratory that directly examine how attention can affect the relationship between neuronal and behavioural performance.

2. THE MAGNITUDE OF MODULATION OF NEURONAL RESPONSES BY ATTENTION

Spatial attention has been found to affect neuronal responses in every visual cortical area examined. Most neurons respond more strongly when the subject attends to a stimulus within their receptive field, compared with attention to another stimulus far from the receptive field. However, the amount by which attention alters neuronal responses can vary greatly between different visual areas, and between neurons within a given visual area. Task demands can also affect the amount by which attention alters responses.

Within any visual area, neurons differ in the degree to which they are influenced by spatial attention. Figure 1 shows a representative distribution of attentional modulations, based on a sample of neurons from area V4 in monkey visual cortex. The average modulation was a 26% increase in response when attention was directed to the receptive field stimulus, but the responses of some neurons were twofold or threefold stronger. The responses of other neurons were not obviously affected by attention, and still others responded more strongly when attention was directed away from the receptive field. A broad range of effects that includes both positive and negative modulations is typical for studies of attention in visual cortex. Why different neurons show different degrees of attentional modulation is not understood, nor has the degree

* Author for correspondence (maunsell@bcm.tmc.edu).

One contribution of 14 to a Discussion Meeting Issue 'The physiology of cognitive processes'.

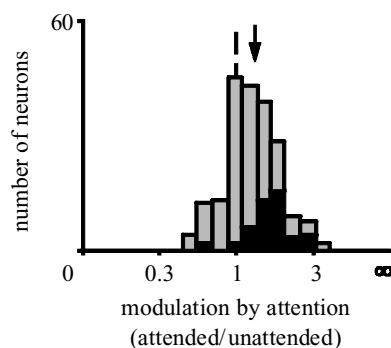


Figure 1. Typical distribution of attentional modulation for visual cortex. Responses were recorded from 197 neurons in area V4 while monkeys performed a task that directed their attention towards or away from a stimulus in the receptive field of the neuron being recorded. Most neurons responded more strongly when the animal paid attention to the stimulus in the receptive field (average increase 26%, indicated by arrow). However, the distribution was broad, and some neurons responded more strongly when the animal directed its attention away from the receptive field. Response changes that were statistically significant are shown in black. (Data from McAdams & Maunsell (1999a).)

of attentional modulation been shown to correlate with other properties of the neurons.

Attentional modulation also differs between visual areas. Modulation by attention is typically weakest in the earliest stages of visual cortex, and strongest in the latest stages. This increase has not been studied extensively, but it is apparent in comparing results from different reports. The best data come from individual studies that have recorded from different cortical areas in the same animals while they performed a given task. Figure 2 shows the average attentional modulation in different cortical areas that were examined in this way in our laboratory. Average attentional modulation is plotted as a function of the level of cortical processing, as defined by the hierarchy of Felleman & Van Essen (1991). In each study, areas at later stages had stronger average attentional modulation. These results were based on extracellular recording of individual action potentials, but stronger modulation at later stages of visual cortex has also been seen by using current source density measurements (Mehta *et al.* 2000).

It is not known why attentional modulations are stronger in later cortical levels. While it is tempting to imagine that neuronal activity related to attention accumulates in successive levels of processing, this seems unlikely given that sensory responses arriving from different sources do not appear to accumulate in this way. Similarly, the notion that neuronal signals related to attention are inserted at the latest stages of processing and diminish as they are fed back to earlier stages seems unlikely. It seems probable that the degree of modulation is optimized for each cortical level. There is little reason to believe that the cerebral cortex would be unable to prevent undesirable accumulation or diminution of signals that altered sensory responses.

In addition to the range of attentional modulation seen within and between visual areas, individual neurons can show different degrees of attentional modulation depending on task demands. Task demand can be affected either by the number of relevant items or by the com-

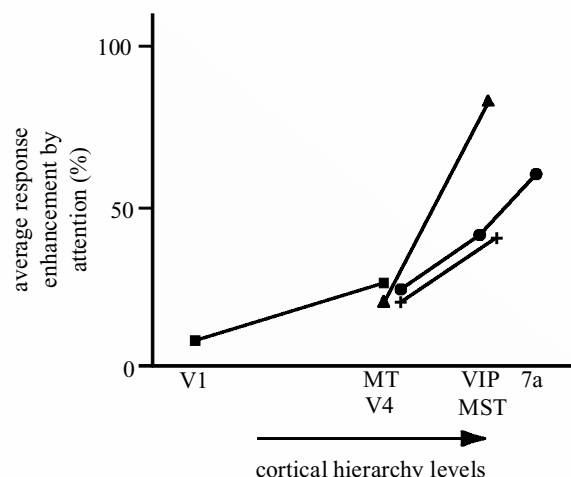


Figure 2. Average attentional modulation in visual cortical areas. The average response enhancement reported from studies that measured the effects of attention in two or more cortical areas in the same subjects while they performed a given task are shown. Positions on the x-axis are assigned according to the hierarchical levels defined by Felleman & Van Essen (1991). More attention modulation is found in later stages of cortical processing (squares, McAdams & Maunsell (1999a); crosses, Treue & Maunsell (1999); circles, Ferrera *et al.* (1994); triangles, Cook & Maunsell (2002)). (Based on Cook & Maunsell (2002).)

plexity of the processing to be performed on the relevant items (Lavie & Tsal 1994; Urbach & Spitzer 1995; Sade & Spitzer 1998). Mountcastle *et al.* (1981, 1987) showed that most neurons in area 7a and neighbouring regions were far more responsive when an animal was engaged in a visual task than when the same retinal stimuli were presented during periods of alert wakefulness when no fixation was required. Neuronal responses in inferotemporal cortex have also been found to be progressively stronger as a fixating animal goes from a situation where the stimulus is irrelevant, to monitoring the stimulus to detect its dimming, to discriminating the shape or texture of the stimulus (Spitzer & Richmond 1991). Similarly, the responses of V4 neurons are stronger when an animal performs a more difficult version of an orientation change detection task (Spitzer *et al.* 1988). The effects of task demand may be related to arousal or vigilance, but they can affect spatial attention. In some circumstances, modulation by spatial attention can be more than twice as strong during a difficult task (Boudreau & Maunsell 2001).

The attentional modulation of the responses of individual neurons can vary not only between tasks but also within trials. Motter (1994) provided a clear example of the dynamics of attention. He trained monkeys to do a task in which an instruction redirected their spatial attention in the middle of some trials. Neurons in V4 changed their responses within a few hundred milliseconds of the new instruction. Other studies have shown that attentional modulation can vary systematically during the course of task trials (e.g. McAdams & Maunsell 1999a; Reynolds *et al.* 2000). These observations indicate that attentional modulation of individual neurons is dynamic, and varies over a time-course of no more than a few hundred milliseconds.

In summary, attentional modulation of neuronal signals

varies substantially within and between areas, and the modulation of individual neurons can vary depending on task demands, and probably other factors. This variable aspect of attentional modulations makes comparisons between different experiments problematic. It also means that there is little point in seeking a specific, fundamental value to describe the effect of attention on a neuron or area.

Although attention substantially alters sensory responses (e.g. figures 1 and 2), it does not usually eliminate responses to unattended stimuli. Occasional neurons are described that respond only when the animal attends to the stimulus in their receptive field, but no report has described this as typical for neurons in a visual area. While it is likely that more neurons would show this sort of gating by attention in extremely demanding tasks, those tasks are not representative of the effort of most visual tasks. When animals perform tasks that approximate the attentional demands of everyday visual functions, attention substantially emphasizes the cortical representation of behaviourally relevant stimuli relative to others, but does not obliterate most signals about other stimuli.

3. THE QUALITY OF MODULATIONS OF NEURONAL RESPONSES BY ATTENTION

In addition to changing the magnitude of neuronal responses, attention might affect the information contained in the neuronal signal. For example, does attention change the preferred orientation or speed, or the sharpness of tuning for different stimulus attributes? The relationship between tuning for stimulus dimensions and the quality of information conveyed by a population of neurons is complex (Zhang & Sejnowski 1999), and it is difficult to argue that attention should have a particular effect on selectivity to produce better behavioural performance. Nevertheless, it is important to examine whether changes in neuronal stimulus selectivity are associated with attention. Only a handful of studies have addressed this question, but they suggest that attention does not systematically alter the selectivity of neurons. Instead, attention appears to act by increasing the responsiveness of neurons without changing the selectivity or reliability of responses.

McAdams & Maunsell (1999a) measured the orientation tuning of V4 neurons when attention was directed towards the orientation of a stimulus in their receptive fields, and when it was directed to the colour of a distant stimulus. Responses were stronger with attention directed to the receptive field, but responses to all orientations were increased by the same proportion. Figure 3 shows average orientation tuning curves for the two conditions in that study. Although responses were on average *ca.* 30% stronger when attention was directed to the receptive field, there was no systematic change in the sharpness of orientation tuning. Treue & Martínez Trujillo (1999) similarly found that attention increased the strength of responses of neurons in MT but did not change the sharpness of their directional tuning. In V4, the spatial profiles of receptive fields appear to scale proportionally when attention is directed at different locations near the receptive field, without changing the shape of the receptive field profile (Connor *et al.* 1996, 1997). Additionally, orientation tuning curves measured in inferotemporal cortex

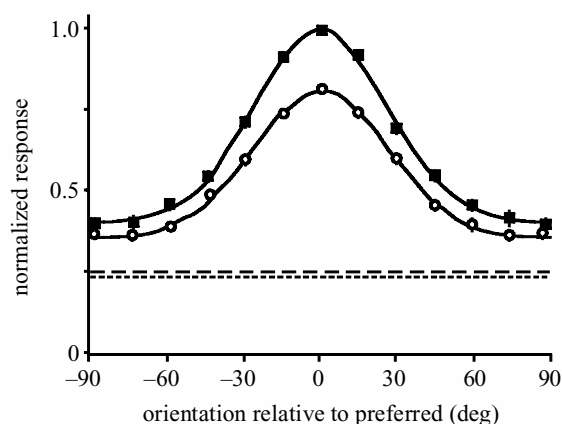


Figure 3. Average normalized tuning curves for 262 V4 neurons tested when the stimulus was attended (filled symbols) or ignored (open symbols). Tuning curves for each cell were normalized to the peak response in the attended condition, and shifted left or right to bring each cell's preferred orientation into alignment. Gaussian functions were then fitted to the averaged data for each behavioural state. Dashed and dotted lines represent the average spontaneous activities in the attended and unattended states respectively. Although the amplitude of the normalized tuning curves changed markedly when the animal attended to the stimulus (amplitude: 0.46–0.60), the width did not ($\sigma = 37.8$ – 37.5°). (Data from McAdams & Maunsell (1999a).)

using reward-contingent and non-reward-contingent stimuli differ in response strength but not in the sharpness of tuning (Vogels & Orban 1994).

Similarly, attention does not appear to alter the underlying reliability of cortical responses. When the same stimulus is presented many times, cortical neurons respond with different numbers of spikes for different presentations. This variability is captured by the relationship between the mean strength of a neuron's response and the variance of the response measured over many presentations of the same stimulus. In most cases, the variance of the spike counts is proportional, and generally close to, the mean spike count (e.g. Tolhurst *et al.* 1981; Geisler & Albrecht 1997). Attention does not affect the relationship between the mean and variance of responses for neurons in area V4 (McAdams & Maunsell 1999b). Although attention changes the average responsiveness of neurons, it does not alter the variance associated with a given level of response.

While in many situations attention appears to increase the sensitivity of neurons without obviously affecting their underlying stimulus selectivity, there are cases where attention does alter selectivity. For example, Moran & Desimone (1985) described how attention altered the receptive fields of V4 neurons to emphasize the region around the stimulus to which the subject was attending (see also Luck *et al.* 1997). A shift in a neuron's receptive field profile is not consistent with a simple change in its sensitivity. Nevertheless, a shift in receptive fields can be explained by simple changes in sensitivity that occur at earlier levels in cortical processing (see Maunsell & McAdams 2001), and are consistent with a mechanism that increases responses to attended stimuli relative to others in the scene.

While more stimulus dimensions and more visual areas

must be examined before firm conclusions can be drawn, the results described above indicate that attention selectively enhances responses to attended stimuli, without changing the feature selectivity of the responsive neurons. This enhancement is similar to what occurs when neurons are presented with stimuli that are more intense, or better suited to the stimulus preferences of the neuron, such as better matched to a preferred colour or speed of motion. For example, increasing stimulus contrast typically makes a neuron's response stronger without changing its selectivity for orientation, direction, spatial frequency or stimulus position (Dean 1981; Holub & Morton-Gibson 1981; Sclar & Freeman 1982; Albrecht & Hamilton 1982; Skottun *et al.* 1987; Geisler & Albrecht 1997). Similarly, selectivity for one stimulus dimension generally does not differ greatly between measurements made using a stimulus that is optimal or suboptimal for another stimulus dimension (see McAdams & Maunsell 1999a); e.g. measurements of direction selectivity at an optimal and suboptimal speed, (Rodman & Albright 1987), although there are exceptions (e.g. Roy & Wurtz 1990). Attention also mimics the way that stimulus changes alter responses without affecting the relationship between response strength and response variance (Dean 1981; Tolhurst *et al.* 1983; Snowden *et al.* 1992).

The notion that attention acts in a manner similar to increasing stimulus intensity is supported by the recent results of Reynolds *et al.* (2000). They measured responses of V4 neurons to stimuli of different contrasts, which were or were not attended by the subject. Attention increased neuronal responses in a manner that was similar to the effect of increasing the contrast of each stimulus by a given amount.

4. ATTENTION, NEURONAL RESPONSES AND BEHAVIOURAL PERFORMANCE

Attention does not appear to alter the nature or structure of the representations in cortex. Instead it serves to highlight particular stimuli within the existing representational framework. This observation suggests an explanation for the behavioural effects of selective visual attention. Focusing attention on a stimulus influences neuronal responses in a way that is similar to increasing stimulus intensity. Like attention, increased stimulus intensity results in neurons responding more strongly, but does not sharpen their tuning for different stimulus features. If the effect of attention on neurons was functionally equivalent to higher intensity or salience of the attended stimuli, the behavioural advantages associated with selective attention—better detection, discrimination and reaction time—might be explained by changes in the representations in sensory cerebral cortex. Thus, attention may make attended stimuli appear more intense or salient by enhancing their representations in sensory cortex.

The possibility that changes in neuronal responses in sensory cortex can fully account for the behavioural advantages associated with attention raises an important question about the relationship between neuronal sensitivity and behavioural capabilities. As mentioned in § 1, studies that have compared neuronal and behavioural performance have often found that under appropriate conditions, the signals of individual neurons can be as reliable as the subject's behaviour in distinguishing between stim-

uli (see Parker & Newsome 1998). Does the correspondence between neuronal and behavioural performance remain fixed when attention alters neuronal responses? We examined this question by comparing neuronal and behavioural responses between different attentional states (Cook & Maunsell 2002).

Because many studies of neuronal performance in visual cortex have been done in areas with directionally selective neurons, we trained two monkeys (*Macaca mulatta*) to do a motion detection task that made it possible to measure neuronal and behavioural responses to stimuli in conditions of high and low attention. The animal viewed a screen on which two patches of dynamic random dots appeared. At the start of each trial, there was no net motion in either patch of dots (0% motion coherence), and the animal's task was to rapidly release a lever when dots of either patch began to move coherently in one direction. The motion signal on each trial was randomly selected from preset levels of coherence. One of the patches filled the receptive field of the neuron being recorded, and coherent motion in either patch matched the preferred direction and speed of the neuron.

Motion appeared in only one patch on each trial, and the animal was cued at the start of the trial as to which patch was likely to contain the motion. Critically, the cue was valid on only about 80% of the trials. On the remaining trials motion occurred at an uncued location. These invalidly cued trials made it possible to measure behavioural and neuronal responses to stimuli to which the animal directed relatively little attention (Posner 1980). The behavioural performance of both animals showed that they used the cue on each trial to direct most of their attention to the patch of dots that was likely to contain the motion. Moderate motion coherence was typically detected on more than 80% of trials when it appeared in the cued location, but on fewer than 40% of trials when it appeared in the uncued location. (These behavioural data cannot distinguish whether the animal simultaneously attended to the cued and uncued locations with different amounts of attention allocated to each, or if instead all of its attention was assigned to one patch at any instant, with the proportion of time spent attending to each patch depending on the cue. However, the neurophysiological recordings indicate that the animal allocated different amounts of attention to the two patches, and that these amounts did not vary during the stimulus presentations (see Cook & Maunsell 2002, figure 10).)

We recorded from neurons in MT and VIP. We chose these areas because both contain neurons that respond selectively to motion. Responses from a typical MT neuron are shown in figure 4. The filled symbols in figure 4a show the average response of the neuron to different levels of motion coherence in the receptive field when the animal had been cued to attend to that location. The dashed line shows the average level of response to the 0% motion stimulus that preceded the motion. As expected, stronger motion produced stronger responses. The open symbol shows the neuron's average response to the medium motion strength (15% coherence) in the receptive field on trials in which the animal had been instructed to attend to the other location. Although the same motion appeared in the receptive field, the response was weaker. While measurements of responses to unattended stimuli of dif-

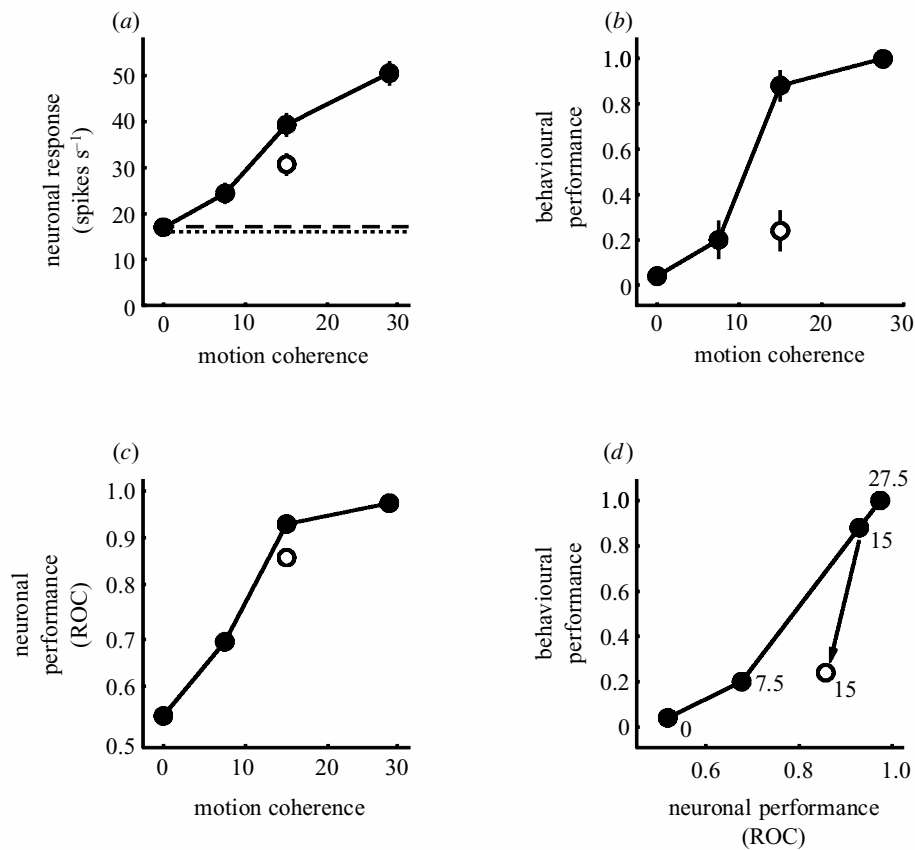


Figure 4. Responses of a representative MT neuron. (a) Responses to motion stimuli with different levels of coherence. Motion coherence corresponds to the strength of the motion signal. At 0% coherence, there is no net motion. Each filled circle is the average response during the first 300 ms after the onset of motion with the coherence indicated. The dashed line is the average response to the 0% coherence stimulus that appeared before the onset of coherent motion. The open circle is the response to 15% coherent motion in the receptive field that appeared on trials when the animal's attention had been directed away from the receptive field. The dotted line is the response to 0% coherent motion in those trials. Error bars in all panels are the standard errors of the means, and are sometimes smaller than the symbols. As expected, directing attention away from the receptive field reduced neuronal responses. (b) Behavioural performance for the same trials used to collect the neuronal responses. The open circle shows that the animal's ability to detect the 15% coherent stimulus was greatly reduced when its attention was directed away from the location of the motion. (c) Neuronal performance. Neuronal responses were converted into performance using ROC based on 300 ms immediately before and after the onset of motion. (d) The data from (b and c) are replotted parameterized by motion coherence. The value by each symbol gives the associated motion coherence. The data from the cases when the animal attended to the location of the motion define a function, but the open symbol, from trials on which the animal's attention was directed away from the stimulus, does not lie on that function. Thus, directing attention away from the stimulus does not have the same effect as reducing motion coherence while keeping attention fixed.

ferent motion coherences would have been valuable, it was not practical to collect those data because invalid cues had to be restricted to a small percentage of the trials.

Figure 4b shows the animal's behavioural performance on the same trials on which the neuronal responses were recorded. The filled symbols plot performance when the animal was attending to the receptive field and the motion occurred there. The open symbol is the behavioural performance for detecting the onset of the medium motion coherence in the receptive field when attention had been directed to the other patch of random dots. Giving the animal an invalid cue caused performance to drop from *ca.* 85% correct to only *ca.* 25% correct.

The question we wanted to address was whether the decreases in neuronal and behavioural performance caused by directing attention were the same as those that might have occurred by keeping attention fixed and reducing the motion strength: is behavioural performance the same for an attended, weak stimulus and an unattended,

strong stimulus if they both produce the same neuronal response? To examine this, we converted the neuronal responses into a measure of performance using ROC analysis (Green & Swets 1966). This analysis evaluates how well an ideal observer could detect the presence of motion if the only signal available were the spikes of the neuron. For this analysis, we counted spikes in 300 ms windows immediately before and after the onset of the motion, for each of the conditions. The ROC performance based on the neuronal signal is plotted in figure 4c as a function of motion coherence. ROC values run from 1.0, which corresponds to perfect performance, to 0.5, which corresponds to performance that is no better than guessing. The neuronal response to the medium motion coherence could support good performance when the receptive field location had been cued, and the neuronal performance was only slightly reduced when the same stimulus appeared with attention directed elsewhere (open symbol).

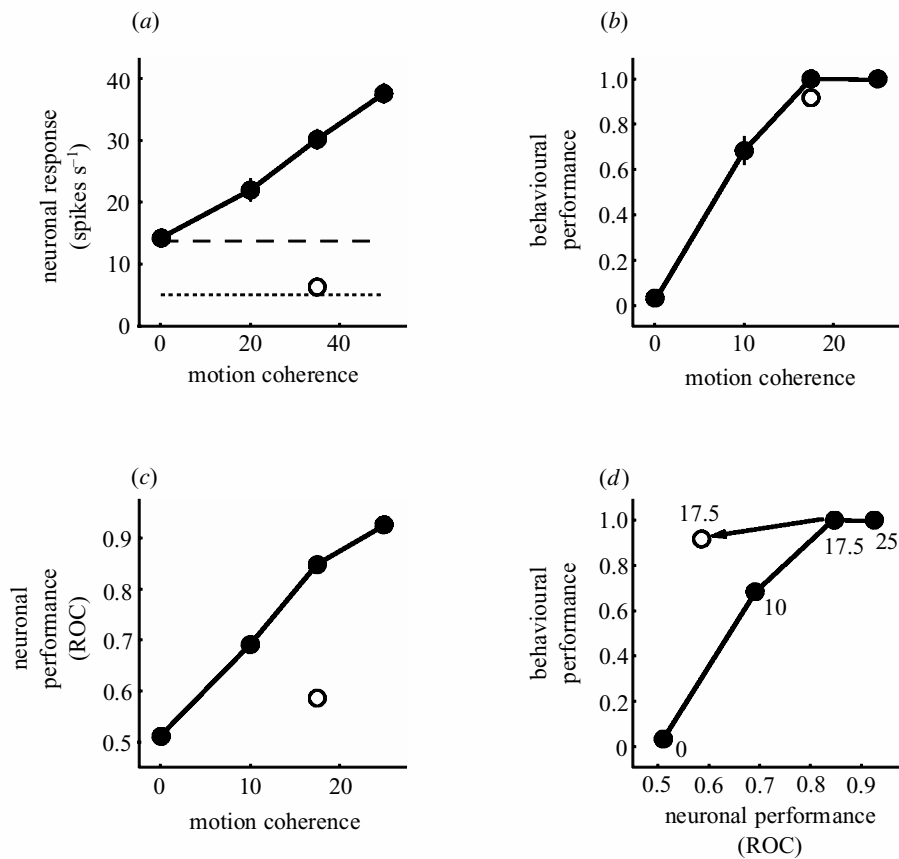


Figure 5. Responses of a representative VIP neuron. Same format as figure 4. (a) Responses to motion stimuli with different levels of coherence; (b) behavioural performance for the same trials used to collect the neuronal responses; (c) neuronal performance; and (d) the data from (b) and (c) replotted parameterized by motion coherence. Attention modulated the neuronal response far more than expected based on its modulation of behavioural performance.

One way to assess whether shifting attention is equivalent to changing stimulus strength is to plot neuronal and behavioural performance against one another, parameterized by the coherence of the motion. In figure 4d, each point represents the behavioural and neuronal performance for a particular motion coherence, as indicated by the numbers beside each point. If the effect of attention were equivalent to changing stimulus strength, then the point for the invalid cue should lie on the function described by the filled symbols. In this case, it does not: it lies substantially to the right of the function. Shifting attention modulated the neuronal response by less than would be needed to account for the modulation of behavioural performance. This result was typical for neurons in MT. The responses of most MT neurons were weakly affected by attention. The average modulation among 93 neurons was about a 15% change in response. Attention had a more pronounced effect on behavioural performance, and for most neurons the attentional modulation of the neuronal response was too small to account for the attentional modulation of the behavioural response.

A different result was found in VIP. VIP receives input from MT (Maunsell & Van Essen 1983; Ungerleider & Desimone 1986), and like MT, most of its neurons are direction selective (Colby *et al.* 1993; Cook & Maunsell 2002). Most neurons in VIP had strong attentional modulation; too strong to match the behavioural modulation by attention. The responses of a representative VIP neuron are shown in figure 5, which has the same format as figure

4. Figure 5a shows that the neuron's responses were profoundly affected by attention. Although the neuron responded strongly to the medium motion coherence when the animal was attending to the receptive field, the same stimulus produced almost no response when the animal was attending elsewhere. The dashed and dotted lines show that attention also affected the response to the 0% motion. The response to unattended moderate motion was less than the response to the 0% coherent motion when it was attended. Although the neuron's responses were devastated when attention was directed elsewhere, in this particular case the animal's behaviour was only slightly reduced (figure 5b). Plotting neuronal performance against behavioural performance showed that attention did not have the same effect as changes in the motion coherence of attended stimuli (figure 5d). Unlike MT, however, the point for the unattended stimulus lies to the left of the function defined by the points for the attended stimuli. This means that the attentional modulation of the neuronal response was stronger than what would be expected based on the modulation of the behavioural response. The responses of most neurons in VIP were strongly modulated by attention, and it was common to find that the neuronal modulation was greater than the associated behavioural modulation.

The responses of all the neurons recorded in MT and VIP are summarized in figure 6. Data from each animal and each area are plotted separately. Each plot shows the average neuronal and behavioural performance associated

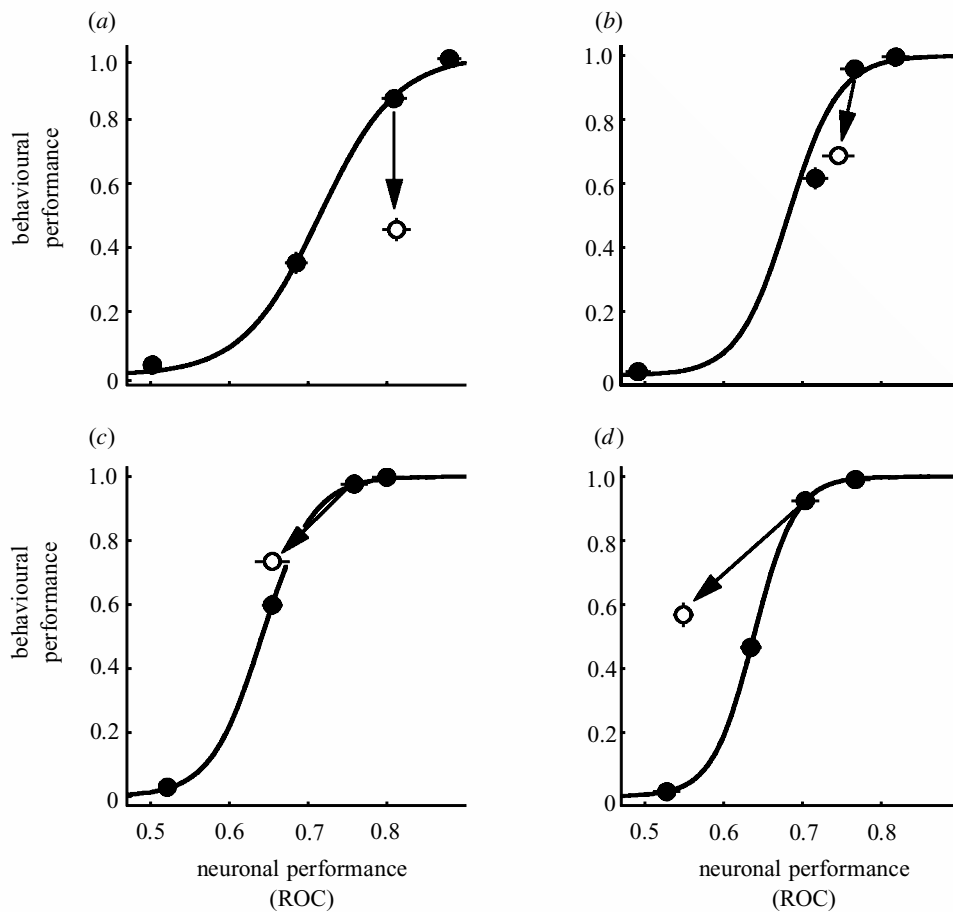


Figure 6. Summary of the effects of attention on neuronal and behavioural performance. (a) Animal 1, MT; (b) animal 2, MT; (c) animal 1, VIP; and (d) animal 2, VIP. Each panel is a plot of neuronal versus behavioural performance in the form shown in figures 4*d* and 5*d*. The filled circles are the average performance for high, medium and low motion coherence and solid bars are sigmoidal fits. Averages from individual neurons were combined to produce these plots. Error bars are the standard errors of the means. Open circles are performance for the medium coherence when the subject's attention was directed away from the stimulus. Although the subjects differed somewhat, the unattended points for the MT are both to the right of the functions for the attended stimuli, and the unattended points for the VIP are both to the left of the functions for the attended stimuli. The effects of attention are not readily equated to changes in motion coherence.

with high, medium and low levels of motion coherence, and for the medium motion coherence when it was in the uncued location. For MT data from both animals, the means for the performance on the uncued condition lie to the right of the function defined by the cued condition, although by different amounts for the two animals. For VIP, the means for the performance on the uncued condition lie to the left of the function, although barely so for animal 1. The differences between the two animals arise in part from greater overall attentional modulation for animal 2. The average neuronal modulation for animal 2 was 28% for MT and 94% for VIP, compared with 12% for MT and 68% for VIP in animal 1.

The analyses in figures 4–6 compare neuronal performance measured using ROC against behavioural performance measured as a percentage of stimuli correctly detected. Might different measures reveal a better correspondence between neuronal and behavioural, or better correspondence between the animals? For example, neuronal response might be measured as absolute or relative rate of firing, and behavioural response might be measured as reaction time. We have tested these and other measures of neuronal and behavioural performance. Different analyses change the shapes of the functions and shift the

unattended points slightly left or right, but none changed the qualitative appearance of the results shown in figure 6 (see Cook & Maunsell 2002). We do not believe that any analysis could bring all the unattended points onto the functions defined by the attended points.

These results show that the relationship between neuronal response and behavioural performance can be changed by attention. Although attention substantially affected behavioural performance, it had relatively little effect on the responsiveness of MT neurons. Thus, similar levels of neuronal performance in MT were associated with different levels of behavioural performance. In VIP, the relationship between neuronal response and performance was also affected by attention, but in this case the attention affected neuronal performance more than behavioural performance.

While there are several possible reasons for this outcome, differences in the amount of attentional modulation across visual cortex provide a probable explanation. Attentional modulation of neuronal responses is greater in later stages of cortical processing (figure 2). Although the origin of this difference is not known, it should have consequences for the relationship between neuronal and behavioural performance. If stimulus–response functions are

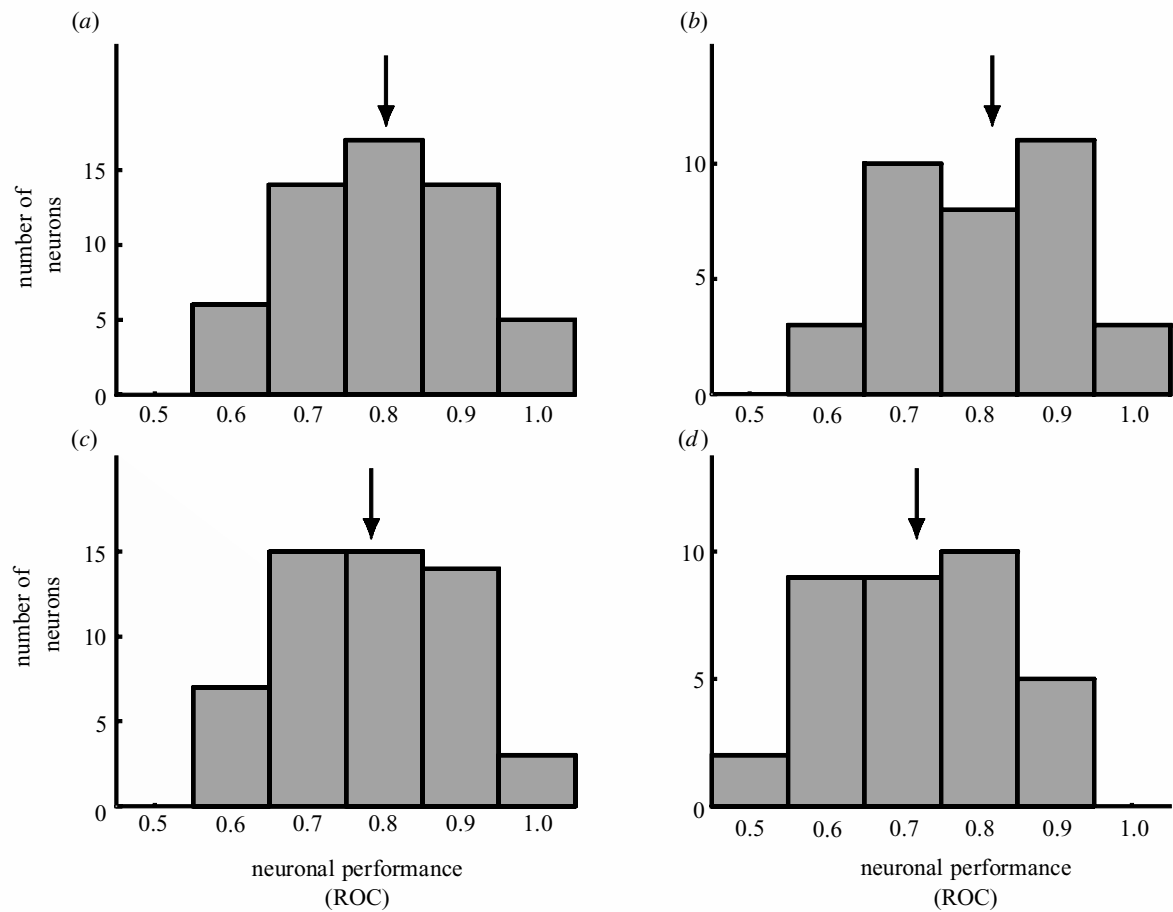


Figure 7. Neuronal performance detecting the onset of stimulus motion. (a) Animal 1, MT; (b) animal 2, MT; (c) animal 1, VIP; and (d) animal 2, VIP. ROC analysis was carried out using 300 ms periods immediately before and after the onset of medium motion coherence when it occurred in the cued location. The median values for the MT were 0.80 and 0.82 for animals 1 and 2, and the median values for the VIP were 0.78 and 0.72 respectively. Thus, the performance of MT neurons was slightly superior. The MT performance was also better for low and high motion coherence.

similar between areas (as they are for MT and VIP; figures 4a and 5a), then a given change in behavioural performance caused by changes in attention will be matched only by a specific change in neuronal response. Because attentional modulation varies between areas, a match between the change in neuronal response and behaviour can be centred only on a particular level of cortical processing. It is possible that the neurons at some level between MT and VIP had an average attentional modulation that was well matched to the average behavioural modulation. If so, then neurons in this region might have responses that closely matched behavioural performance across manipulations of either stimulus strength or attentional state.

The notion that behavioural performance might be most closely associated with neurons that are intermediate in cortical processing raises the question of why behaviours would not be based on the neurons with the greatest attentional modulation. The answer is probably that the neurons with the greatest attentional modulation do not have response properties that are optimal for the task. Neurons with the greatest attentional modulation typically lie at the highest levels of cortical processing, and have complex response properties. For example, although neurons in both MT and VIP are directionally selective, VIP neurons have more elaborate response properties. VIP neurons respond to several types of visual and extraretinal

signals, including optic flow patterns, tactile stimulation of the face, vestibular stimulation, and the position of the eyes in the orbits (Schaafsma & Duysens 1996; Schaafsma *et al.* 1997; Colby & Goldberg 1999). Although these neurons are strongly modulated by attention, their specialization for these more complex response properties is likely to make them less reliable than other neurons for detecting the onset of the motion used in the experiments described here.

An analysis of the responses we collected from MT and VIP indicates that neurons in VIP are not better suited than MT neurons for detecting the motion used in our task. Figure 7 shows distributions of neuronal performance for the two areas. A broad range of performance was found in both areas, but in both animals the median performance was slightly better for MT. These distributions are based on responses to the medium motion coherence when it appeared in the cued location, but MT performance was also better for the other levels of motion coherence. Thus, neurons at later levels of cortical processing are not necessarily better suited for all tasks.

We know of no data that strongly suggest that behavioural performance should always depend on neurons in the highest levels of visual cortex. Basing performance on those representations would introduce a substantial restriction, because far fewer neurons would be available

if only the highest levels are accessed. Rather than depending on those neurons that are most modulated by attention, behavioural performance is likely to be based on those neurons with response properties that provide the most sensitive signals for the current task. This suggestion is consistent with observations on perceptual learning that indicate that earlier or later cortical levels are involved in learning different types of visual discriminations (Ahissar & Hochstein 1997). In the case of our motion detection task, neurons with the greatest sensitivity might be found at a level intermediate to MT and VIP, and behaviours might be most closely linked to the performance of those neurons. Were that the case, it would be natural for behaviour to be modulated by attention to the same extent that those best-suited neurons were modulated by attention. If, in a different task, behavioural performance was best served by representations in later stages of visual cortex, where attentional modulation is strong, we would expect that behaviour would be more susceptible to attentional modulation.

5. CONCLUSION

More data will be needed before we will know whether the relationship between neuronal and behavioural performance persists across attentional states for any region of the visual cortex. Nevertheless, the results described here indicate that a correspondence across attention states will be limited to a restricted level of processing, the location of which will vary depending on the task being performed.

Even if neurons at a given level do preserve a fixed relationship between neuronal and behavioural performance across attentional states for a task, it would not necessarily mean that those neurons were more or less important than others for performing the task. Similarly, it would say nothing about how large a pool of neurons contribute to a decision. A cortical region that had a close match would be a good candidate for being the centre of the relevant neurons, but the pool size might be large or small without appreciably affecting the mean neuronal performance. Additionally, a correspondence between average neuronal performance and average behavioural performance would not reveal which cortical region was most involved, because behavioural performance might be based on the best neurons in a population, rather than the average (see Parker & Newsome 1998).

Much remains to be learned about how the activity of individual neurons is translated into behavioural decisions. Nevertheless, progress in this area is essential for understanding how the brain works. Manipulation of attention may prove to be a valuable approach to exploring how neurons contribute to particular behaviours.

We thank our many colleagues who contributed to the studies reviewed here and provided comments on the manuscript, including W. Bosking, C. E. Boudreau, J. J. DiCarlo, G. M. Ghose, C. Hocker, C. J. McAdams, D. Murray, T. Williford and T. Yang. Supported by NIH R01 EY05911. J.H.R.M. is an investigator with the Howard Hughes Medical Institute.

REFERENCES

- Ahissar, M. & Hochstein, S. 1997 Task difficulty and the specificity of perceptual learning. *Nature* **387**, 401–406.
- Albrecht, D. G. & Hamilton, D. B. 1982 Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.* **48**, 217–237.
- Barlow, H. B. 1985 The twelfth Bartlett memorial lecture: the role of single neurons in the psychology of perception. *Q. J. Exp. Psychol.* **37**, 121–145.
- Boudreau, C. E. & Maunsell, J. H. R. 2001 Is spatial attention a limited resource for V4 neurons? *Soc. Neurosci. Abstr.* **27**, 574.3.
- Braun, J., Koch, C. & Davis, J. L. 2001 *Visual attention and cortical circuits*. Cambridge, MA: MIT.
- Colby, C. L. & Goldberg, M. E. 1999 Space and attention in parietal cortex. *A. Rev. Neurosci.* **22**, 319–349.
- Colby, C. L., Duhamel, J. R. & Goldberg, M. E. 1993 Ventral intraparietal area of the macaque: anatomic location and visual response properties. *J. Neurophysiol.* **69**, 902–914.
- Connor, C. E., Gallant, J. L., Preddie, D. C. & Van Essen, D. C. 1996 Responses in area V4 depend on the spatial relationship between stimulus and attention. *J. Neurophysiol.* **75**, 1306–1308.
- Connor, C. E., Preddie, D. C., Gallant, J. L. & Van Essen, D. C. 1997 Spatial attention effects in macaque area V4. *J. Neurosci.* **17**, 3201–3214.
- Cook, E. P. & Maunsell, J. H. R. 2002 Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. *J. Neurosci.* **22**, 1994–2004.
- Dean, A. F. 1981 The variability of discharge of simple cells in the cat striate cortex. *Exp. Brain Res.* **44**, 437–440.
- Felleman, D. J. & Van Essen, D. C. 1991 Distributed hierarchical processing in the primate cerebral cortex. *Cerebr. Cortex* **1**, 1–47.
- Ferrera, V. P., Rudolph, K. K. & Maunsell, J. H. R. 1994 Responses of neurons in the parietal and temporal visual pathways during a motion task. *J. Neurosci.* **14**, 6171–6186.
- Geisler, W. S. & Albrecht, D. G. 1997 Visual cortex neurons in monkeys and cats: detection, discrimination and identification. *Visual Neurosci.* **14**, 897–919.
- Green, D. M. & Swets, J. A. 1966 *Signal detection theory and psychophysics*. New York: Wiley.
- Holub, R. A. & Morton-Gibson, M. 1981 Response of visual cortical neurons of the cat to moving sinusoidal gratings: response-contrast functions and spatiotemporal interactions. *J. Neurophysiol.* **46**, 1244–1259.
- Lavie, N. & Tsal, Y. 1994 Perceptual load as a major determinant of the locus of selection in visual attention. *Percept. Psychophys.* **56**, 183–197.
- Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. 1997 Neural mechanisms of spatial selective attention in areas V1, V2 and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42.
- McAdams, C. J. & Maunsell, J. H. R. 1999a Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441.
- McAdams, C. J. & Maunsell, J. H. R. 1999b Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron* **23**, 765–773.
- Mehta, A. D., Ulbert, I. & Schroeder, C. E. 2000 Intermodal selective attention in monkeys. I. Distribution and timing of effects across visual areas. *Cerebr. Cortex* **10**, 343–358.
- Maunsell, J. H. R. & McAdams, C. J. 2001 Effects of attention on the responsiveness and selectivity of individual neurons in visual cerebral cortex. In *Visual attention and cortical circuits* (ed. J. Braun, C. Koch & J. L. Davis), pp. 103–119. Cambridge, MA: MIT.

- Maunsell, J. H. R. & Van Essen, D. C. 1983 The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J. Neurosci.* **3**, 2563–2586.
- Moran, J. & Desimone, R. 1985 Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784.
- Motter, B. C. 1994 Neural correlates of feature selective memory and pop-out in extrastriate area V4. *J. Neurosci.* **14**, 2190–2199.
- Mountcastle, V. B., Andersen, R. A. & Motter, B. C. 1981 The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex. *J. Neurosci.* **1**, 1218–1235.
- Mountcastle, V. B., Motter, B. C., Steinmetz, M. A. & Sestokas, A. K. 1987 Common and differential effects of attentive fixation on the excitability of parietal and prestriate (V4) cortical visual neurons in the macaque monkey. *J. Neurosci.* **7**, 2239–2255.
- Parker, A. J. & Newsome, W. T. 1998 Sense and the single neuron: probing the physiology of perception. *A. Rev. Neurosci.* **21**, 227–277.
- Pashler, H. E. 1998 *The psychology of attention*. Cambridge, MA: MIT.
- Posner, M. I. 1980 Orienting of attention. *Q. J. Exp. Psychol.* **32**, 3–25.
- Reynolds, J. H., Pasternak, T. & Desimone, R. 2000 Attention increases sensitivity of V4 neurons. *Neuron* **26**, 703–714.
- Rodman, H. R. & Albright, T. D. 1987 Coding of visual stimulus velocity in area MT of the macaque. *Vision Res.* **27**, 2035–2048.
- Roy, J. P. & Wurtz, R. H. 1990 The role of disparity-sensitive cortical neurons in signalling the direction of self-motion. *Nature* **348**, 160–162.
- Sade, A. & Spitzer, H. 1998 The effects of attentional spread and attentional effort on orientation discrimination. *Spat. Vision* **11**, 367–383.
- Schaafsma, S. J. & Duysens, J. 1996 Neurons in the ventral intraparietal area of awake monkey closely resemble neurons in the dorsal part of the medial superior temporal area in their response to optic flow. *J. Neurophysiol.* **76**, 4056–4068.
- Schaafsma, S. J., Duysens, J. & Gielen, C. C. A. M. 1997 Responses in ventral intraparietal area of awake macaque monkey to optic flow patterns corresponding to rotation of planes in depth can be explained by translation and expansion effects. *Visual Neurosci.* **14**, 633–646.
- Sclar, G. & Freeman, R. D. 1982 Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Exp. Brain Res.* **46**, 457–461.
- Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I. & Freeman, R. D. 1987 The effects of contrast on visual orientation and spatial frequency discrimination: a comparison of single cells and behavior. *J. Neurophysiol.* **57**, 773–786.
- Snowden, R. J., Treue, S. & Andersen, R. A. 1992 The response of neurons in area V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp. Brain Res.* **88**, 389–400.
- Spitzer, H. & Richmond, B. J. 1991 Task difficulty: ignoring, attending to, and discriminating a visual stimulus yield progressively more activity in inferior temporal neurons. *Exp. Brain Res.* **83**, 340–348.
- Spitzer, H., Desimone, R. & Moran, J. 1988 Increased attention enhances both behavioral and neuronal performance. *Science* **240**, 338–340.
- Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. 1981 The dependence of response amplitude and variance of cat visual cortical neurons on stimulus contrast. *Exp. Brain Res.* **41**, 414–419.
- Tolhurst, D. J., Movshon, J. A. & Dean, A. F. 1983 The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**, 775–785.
- Treue, S. & Martínez Trujillo, J. C. 1999 Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* **399**, 575–579.
- Treue, S. & Maunsell, J. H. R. 1999 Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J. Neurosci.* **19**, 7591–7602.
- Ungerleider, L. G. & Desimone, R. 1986 Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* **248**, 190–222.
- Urbach, D. & Spitzer, H. 1995 Attentional effort modulated by task difficulty. *Vision Res.* **35**, 2169–2177.
- Vogels, R. & Orban, G. A. 1994 Activity of inferior temporal neurons during orientation discrimination with successively presented gratings. *J. Neurophysiol.* **71**, 1428–1451.
- Zhang, K. & Sejnowski, T. J. 1999 Neuronal tuning: to sharpen or broaden? *Neural Comput.* **11**, 75–84.

GLOSSARY

MST: medial superior temporal area
 MT: middle temporal visual area
 ROC: receiver operating characteristic
 VIP: ventral intraparietal area

Which Model to Use for Cortical Spiking Neurons?

Eugene M. Izhikevich

Abstract—We discuss the biological plausibility and computational efficiency of some of the most useful models of spiking and bursting neurons. We compare their applicability to large-scale simulations of cortical neural networks.

Index Terms—Chaos, Hodgkin–Huxley, pulse-coupled neural network (PCNN), quadratic integrate-and-fire (I&F), spike-timing.

I. INTRODUCTION

DURING last few years we have witnessed a shift of the emphasis in the artificial neural network community toward spiking neural networks. Motivated by biological discoveries, many studies (see this volume) consider pulse-coupled neural networks with spike-timing as an essential component in information processing by the brain.

In any study of network dynamics, there are two crucial issues which are: 1) what model describes spiking dynamics of each neuron and 2) how the neurons are connected. Inappropriate choice of the spiking model or the connectivity may lead to results having nothing to do with the information processing by the brain. In this paper, we consider the first issue, i.e., we compare and contrast various models of spiking neurons.

In Section II and Fig. 1, we review important neuro-computational features of real neurons and their contribution to temporal coding and spike-timing information processing. In Section III, we consider various models of spiking neurons and rank them according to: 1) the number of neuro-computational features they can reproduce; and 2) their implementation efficiency, i.e., the number of floating-point operations (addition, multiplication, etc.) needed to simulate the model during a 1 ms time span. The results of our comparison are summarized in Fig. 2. We compare the utility of the models to large-scale simulations of cortical networks in Section IV.

II. NEURO-COMPUTATIONAL FEATURES

In Fig. 1, we review 20 of the most prominent features of biological spiking neurons. The goal of this section is to illustrate the richness and complexity of spiking behavior of individual neurons in response to simple pulses of dc current. What happens when only tens (let alone billions) of such neurons are coupled together is beyond our comprehension. Using some of the models discussed in Section III, the reader can simulate thousands of cortical neurons in real time with 1 ms resolution.

Manuscript received May 17, 2003; revised June 31, 2004. This work was supported in part by the Neurosciences Research Foundation and in part by a Grant from the Alafi Family Foundation.

The author is with the The Neurosciences Institute, San Diego, CA 92121 USA (e-mail: Eugene.Izhikevich@nsi.edu).

Digital Object Identifier 10.1109/TNN.2004.832719

A. Tonic Spiking

Most neurons are excitable, that is, they are quiescent but can fire spikes when stimulated. To test this property, neurophysiologists inject pulses of dc current via an electrode attached to the neuron and record its membrane potential. The input current and the neuronal response are usually plotted one beneath the other, as in Fig. 1(a). While the input is on, the neuron continues to fire a train of spikes. This kind of behavior, called tonic spiking, can be observed in the three types of cortical neurons: regular spiking (RS) excitatory neurons, low-threshold spiking (LTS), and fast spiking (FS) inhibitory neurons [1], [6]. Continuous firing of such neurons indicate that there is a persistent input.

B. Phasic Spiking

A neuron may fire only a single spike at the onset of the input, as in Fig. 1(b), and remain quiescent afterwards. Such a response is called phasic spiking, and it is useful for detection of the beginning of stimulation.

C. Tonic Bursting

Some neurons, such as the chattering neurons in cat neocortex [7], fire periodic bursts of spikes when stimulated, as in Fig. 1(c). The interburst (i.e., between bursts) frequency may be as high as 50 Hz, and it is believed that such neurons contribute to the gamma-frequency oscillations in the brain.

D. Phasic Bursting

Similarly to the phasic spikers, some neurons are phasic bursters, as in Fig. 1(d). Such neurons report the beginning of the stimulation by transmitting a burst.

There are three major hypothesis on the importance of bursts in the brain which are: 1) bursts are needed to overcome the synaptic transmission failure and reduce neuronal noise [20]; 2) Bursts can transmit saliency of the input, because the effect of a burst on the postsynaptic neuron is stronger than the effect of a single spike; and 3) bursts can be used for selective communication between neurons [14], where the interspike frequency within the bursts encodes the channel of communication. A good model of a cortical neuronal network cannot neglect bursting neurons.

E. Mixed Model (Bursting Then Spiking)

Intrinsically bursting (IB) excitatory neurons in mammalian neocortex [1] can exhibit a mixed type of spiking activity depicted in Fig. 1(e). They fire a phasic burst at the onset of stimulation and then switch to the tonic spiking mode. It is not clear what kind of computation such a neuron can do in addition to detecting the onset and reporting the extent of stimulation.

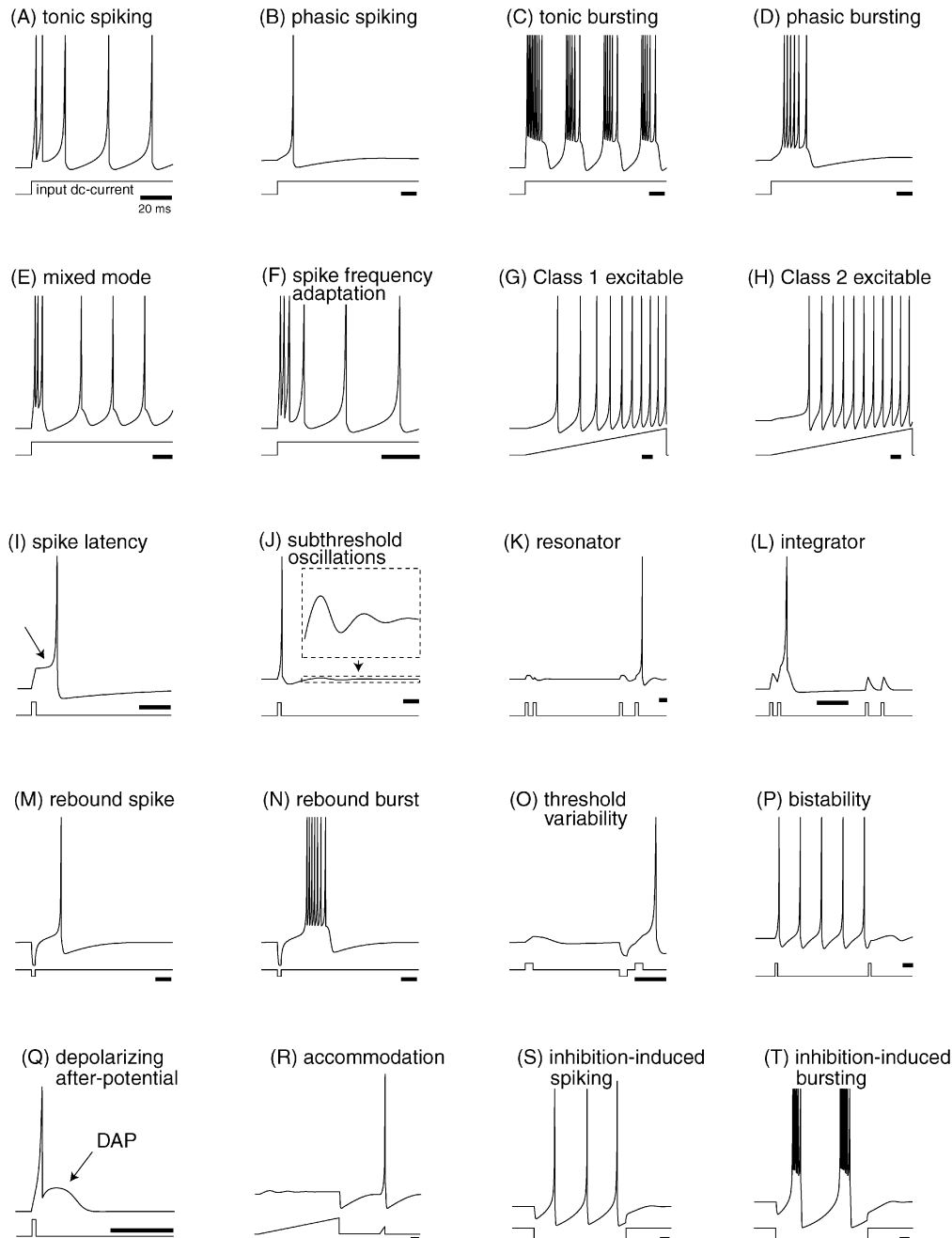


Fig. 1. Summary of the neuro-computational properties of biological spiking neurons. Shown are simulations of the same model (1) and (2), with different choices of parameters. Each horizontal bar denotes a 20-ms time interval. The MATLAB file generating the figure and containing all the parameters, as well as interactive matlab tutorial program can be downloaded from the author's website. This figure is reproduced with permission from www.izhikevich.com. (Electronic version of the figure and reproduction permissions are freely available at www.izhikevich.com).

F. Spike Frequency Adaptation

The most common type of excitatory neuron in mammalian neocortex, namely the regular spiking (RS) cell, fires tonic spikes with decreasing frequency, as in Fig. 1(f). That is, the frequency is relatively high at the onset of stimulation, and then it adapts. Low-threshold spiking (LTS) inhibitory neurons also have this property. The interspike frequency of such cells may encode the time elapsed since the onset of the input.

G. Class 1 Excitability

The frequency of tonic spiking of neocortical RS excitatory neurons depends on the strength of the input, and it may span

the range from 2 Hz to 200 Hz, or even greater. The ability to fire low-frequency spikes when the input is weak (but superthreshold) is called Class 1 excitability [8], [17], [22]. Class 1 excitable neurons can encode the strength of the input into their firing rate, as in Fig. 1(g).

H. Class 2 Excitability

Some neurons cannot fire low-frequency spike trains. That is, they are either quiescent or fire a train of spikes with a certain relatively large frequency, say 40 Hz, as in Fig. 1(h). Such neurons are called Class 2 excitable [8], [17], [22]. Their firing rate is a poor predictor of the strength of stimulation.

I. Spike Latency

Most cortical neurons fire spikes with a delay that depends on the strength of the input signal. For a relatively weak but superthreshold input, the delay, also called spike latency, can be quite large, as in Fig. 1(i). The RS cells in mammalian cortex can have latencies of tens of ms. Such latencies provide a spike-timing mechanism to encode the strength of the input.

J. Subthreshold Oscillations

Practically every brain structure has neurons capable of exhibiting oscillatory potentials, as in Fig. 1(j). The frequency of such oscillations play an important role and such neurons act as bandpass filters, as we discuss next.

K. Frequency Preference and Resonance

Due to the resonance phenomenon, neurons having oscillatory potentials can respond selectively to the inputs having frequency content similar to the frequency of subthreshold oscillations. Such neurons can implement frequency-modulated (FM) interactions and multiplexing of signals [14], [16]. In Fig. 1(k), we stimulate such a neuron with two doublets (pairs of spikes) having different interspike frequencies. The neuron responds only to the doublet whose frequency resonates with the frequency of subthreshold oscillations. Such neurons are called resonators.

L. Integration and Coincidence Detection

Neurons without oscillatory potentials act as integrators. They prefer high-frequency input; the higher the frequency the more likely they fire, as in Fig. 1. This can be useful for detecting coincident or nearly coincident spikes.

M. Rebound Spike

When a neuron receives and then is released from an inhibitory input, it may fire a post-inhibitory (rebound) spike, as in Fig. 1(m). This phenomenon is related to the anodal break excitation in excitable membranes. Many spiking neurons can fire in response to brief inhibitory inputs thereby blurring the difference between excitation and inhibition.

N. Rebound Burst

Some neurons, including the thalamo-cortical cells, may fire post-inhibitory bursts, as in Fig. 1(n). It is believed that such bursts contribute to the sleep oscillations in the thalamo-cortical system.

O. Threshold Variability

A common misconception in the artificial neural network community is the belief that spiking neurons have a fixed voltage threshold. It is well-known that biological neurons have a variable threshold that depends on the prior activity of the neurons. In Fig. 1(o), we first stimulate a neuron with a brief excitatory pulse of current that produces 10 mV depolarization. The neuron does not fire, hence, the input is subthreshold. Then, we apply a brief inhibitory input and then exactly the same “subthreshold” pulse of current. The neuron fires the second time because its “threshold” was lowered by the pre-

ceding inhibitory input. Hence, the same 10-mV depolarization can be subthreshold or superthreshold depending on the prior activity. Interestingly, a preceding excitatory pulse might raise the threshold and make the neuron less excitable.

P. Bistability of Resting and Spiking States

Some neurons can exhibit two stable modes of operation: resting and tonic spiking (or even bursting). An excitatory or inhibitory pulse can switch between the modes, as in Fig. 1(p), thereby creating an interesting possibility for bistability and short-term memory. Notice that to switch from the tonic spiking to resting mode, the input must arrive at an appropriate phase of oscillation, thereby emphasizing the importance of spike-timing in such information processing.

Q. Depolarizing After-Potentials

After firing a spike, the membrane potential of a neuron may exhibit a prolonged after-hyperpolarization (AHP) as, e.g., in Fig. 1(b), I or M, or a prolonged depolarized after-potential (DAP), as in Fig. 1(q). Such DAPs can appear because of dendritic influence, because of a high-threshold inward currents activated during the spike, or because of an interplay between subthreshold voltage-gated currents. In any case, such a neuron has shortened refractory period and it becomes superexcitable.

R. Accommodation

Neurons are extremely sensitive to brief coincident inputs, but may not fire in response to a strong but slowly increasing input, as we illustrate in Fig. 1(r). The slowly ramped current in the figure does not elicit a spike, while a smaller but sharply ramped current elicits a spike. During the slow ramp, the inward currents have enough time to inactivate and outward currents have enough time to activate, so the neuron accommodates, becomes less excitable and cannot generate a spike.

S. Inhibition-Induced Spiking

A bizarre feature of many thalamo-cortical neurons is that they are quiescent when there is no input, but fire when hyperpolarized by an inhibitory input or an injected current, as we illustrate in Fig. 1(s). This happens because the injected current activates the h-current and deinactivates calcium T-current, leading to tonic spiking.

T. Inhibition-Induced Bursting

Instead of spiking, a thalamo-cortical neuron can fire tonic bursts of spikes in response to a prolonged hyperpolarization, as in Fig. 1(t). It is believed that such bursting takes place during spindle wave oscillations in the thalamo-cortical system and it plays an important role in sleep rhythms.

No model should exhibit all these 20 neurocomputational properties simultaneously simply because some of the properties are mutually exclusive. For example, a neuron cannot be an integrator and a resonator at the same time. However, there are models that can easily be tuned to exhibit each such property. For example, all of the neuronal responses in Fig. 1 were obtained using a simple spiking model having four easily tunable parameters [15].

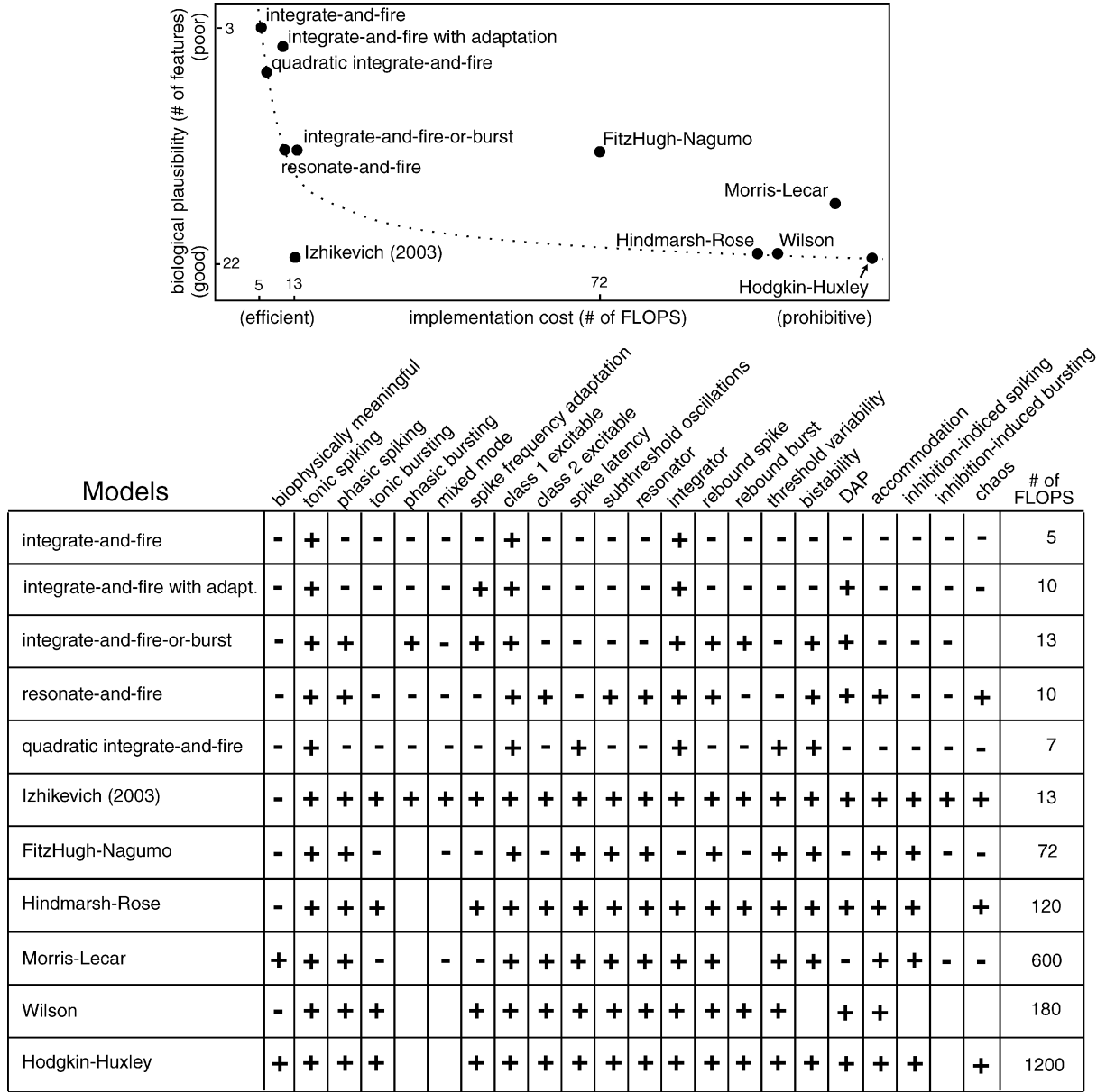


Fig. 2. Comparison of the neuro-computational properties of spiking and bursting models; see Fig. 1. “# of FLOPS” is an approximate number of floating point operations (addition, multiplication, etc.) needed to simulate the model during a 1 ms time span. Each empty square indicates the property that the model should exhibit in principle (in theory) if the parameters are chosen appropriately, but the author failed to find the parameters within a reasonable period of time.

III. SPIKING MODELS

Below we review some widely used models of spiking and bursting neurons that can be expressed in the form of ordinary differential equations (ODE) (thus, we exclude the spike response model [5]). In addition to the 20 neuro-computational features reviewed above, we also consider whether the models have biophysically meaningful and measurable parameters, and whether they can exhibit autonomous chaotic activity. We start with the simplest models first. The summary of our comparison is in Fig. 2.

Throughout this section, v denotes the membrane potential and v' denotes its derivative with respect to time. All the parameters in the models are chosen so that v has mV scale and the time has ms scale. To compare computational cost, we assume that each model, written as a dynamical system $\dot{x} = f(x)$, is

implemented using a fixed-step first-order Euler method $x(t + \tau) = x(t) + \tau f(x(t))$ with the integration time step τ chosen to achieve a reasonable numerical accuracy.

A. I&F

One of the most widely used models in computational neuroscience is the leaky integrate-and-fire (I&F) neuron

$$v' = I + a - bv, \text{ if } v \geq v_{\text{thresh}}, \text{ then } v \leftarrow c$$

where v is the membrane potential, I is the input current, and a, b, c , and v_{thresh} are the parameters. When the membrane potential v reaches the threshold value v_{thresh} , the neuron is said to fire a spike, and v is reset to c .

The I&F neuron is Class 1 excitable; it can fire tonic spikes with constant frequency, and it is an integrator. It is the simplest

model to implement when the integration time step τ is 1 ms. Indeed, the iteration $v(t+1) = v(t) + (I + a - bv(t))$ takes only four floating-point operations (additions, multiplications, etc.) plus one comparison with the threshold v_{thresh} . Because I&F has only one variable, it cannot have phasic spiking, bursting of any kind, rebound responses, threshold variability, bistability of attractors, or autonomous chaotic dynamics. Because of the fixed threshold, the spikes do not have latencies. In summary, despite its simplicity, I&F is one of the worst models to use in simulations, unless one wants to prove analytical results.

B. I&F with Adaptation

The I&F model is one-dimensional (1-D), hence it cannot burst or have other properties of cortical neurons. One may think that having a second linear equation

$$\begin{aligned} v' &= I + a - bv + g(d - v) \\ g' &= \frac{(e\delta(t) - g)}{\tau} \end{aligned}$$

describing activation dynamics of a high-threshold K current can make an improvement, e.g., endow the model with spike-frequency adaptation. Indeed, each firing increases the K activation gate g via Dirac delta function δ and produces an outward current that slows down the frequency of tonic spiking. Simulations of this model take 10 floating point operations/1 ms time step, yet the model still lacks many important properties of cortical spiking neurons.

C. Integrate-and-Fire-or-Burst

Smith and coauthors [24] suggested an improvement—integrate-and-fire-or-burst (I&FB) model

$$\begin{aligned} v' &= I + a - bv + gH(v - v_h)h(v_T - v) \\ \text{if } v &= v_{\text{thresh}}, \text{ then } v \leftarrow c \\ h' &= \begin{cases} \frac{-h}{\tau^-}, & \text{if } v > v_h \\ \frac{(1-h)}{\tau^+}, & \text{if } v < v_h \end{cases} \end{aligned}$$

to model thalamo-cortical neurons. Here h describes the inactivation of the calcium T-current, g , v_h , v_T , τ^+ , and τ^- are parameters describing dynamics of the T-current, and H is the Heaviside step function.

Having this kind of a second variable creates the possibility for bursting and other interesting regimes summarized in Fig. 2. But this comes with a price: It takes between 9 and 13 operations (depending on the value of v) to simulate 1 ms of the model.

D. Resonate-and-Fire

The resonate-and-fire neuron [16] is a two-dimensional (2-D) analogue of the I&F neuron

$$\begin{aligned} z' &= I + (b + \omega)z \\ \text{if } \text{Im } z &= a_{\text{thresh}}, \text{ then } z \leftarrow z_0(z) \end{aligned}$$

where the real part of the complex variable z is the membrane potential. Here b , ω , and a_{thresh} are parameters, and $z_0(z)$ is

an arbitrary function describing activity-dependent after-spike reset. The resonate-and-fire model is simple and efficient—it takes 10 operations to simulate 1 ms. When the frequency of oscillation $\omega = 0$, it becomes an integrator. Its neuro-computational properties are summarized in Fig. 2.

E. Quadratic I&F

An alternative to the leaky I&F neuron is the quadratic I&F neuron, also known as the theta-neuron [2], [3] or the Ermentrout–Kopell canonical model [10] (when it is written in a trigonometric form). We present it here following [19]

$$\begin{aligned} v' &= I + a(v - v_{\text{rest}})(v - v_{\text{thresh}}) \\ \text{if } v &= v_{\text{peak}}, \quad \text{then } v \leftarrow v_{\text{reset}} \end{aligned}$$

where v_{rest} and v_{thresh} are the resting and threshold values of the membrane potential. This model is canonical in the sense that any Class 1 excitable system described by smooth ODEs can be transformed into this form by a continuous change of variables [18]. It takes only seven operations to simulate 1 ms of the model, and this should be the model of choice when one simulates large-scale networks of integrators. Unlike its linear analogue, the quadratic I&F neuron has spike latencies, activity-dependent threshold (which is v_{thresh} only when $I = 0$), and bistability of resting and tonic spiking modes.

F. Spiking Model by Izhikevich (2003)

All of the responses in Fig. 1 were obtained using a simple model of spiking neurons proposed recently by Izhikevich [15]

$$v' = 0.04v^2 + 5v + 140 - u + I \quad (1)$$

$$u' = a(bv - u) \quad (2)$$

with the auxiliary after-spike resetting

$$\text{if } v \geq +30 \text{ mV}, \quad \text{then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d. \end{cases} \quad (3)$$

Here variable v represents the membrane potential of the neuron and u represents a membrane recovery variable, which accounts for the activation of K^+ ionic currents and inactivation of Na^+ ionic currents, and it provides negative feedback to v . After the spike reaches its apex (+30 mV), the membrane voltage and the recovery variable are reset according to the (3). If v skips over 30, then it first is reset to 30, and then to c so that all spikes have equal magnitudes. The part $0.04v^2 + 5v + 140$ is chosen so that v has mV scale and the time has ms scale. Geometrical derivation of the model based on fast and slow nullclines can be found in [11].

The model can exhibit firing patterns of all known types of cortical neurons with the choice of parameters a , b , c , and d given in [15]. It takes only 13 floating point operations to simulate 1 ms of the model, so it is quite efficient in large-scale simulations of cortical networks. When $(a, b, c, d) = (0.2, 2, -56, -16)$ and $I = -99$, the model has chaotic spiking activity, though the integration time step τ should be small to achieve adequate numerical precision.

We stress that +30 mV in (3) is not a threshold, but the peak of the spike. The threshold value of the model neuron is between -70 and -50, and it is dynamic, as in biological neurons. To build intuition and understanding of the dynamics of the model, the reader is advised to download an interactive MATLAB tutorial program from the author's webpage and play with the model and its parameters. In particular, the reader could explore all 20 neuro-computational properties in Fig. 1.

G. FitzHugh–Nagumo

The parameters in the FitzHugh–Nagumo model [4]

$$\begin{aligned} v' &= a + bv + cv^2 + dv^3 - u \\ u' &= \varepsilon(ev - u) \end{aligned}$$

can be tuned so that the model describes spiking dynamics of many resonator neurons. Its neuro-computational properties are summarized in Fig. 2. Since one needs to simulate the shape of each spike, the time step in the model must be relatively small, e.g., $\tau = 0.25$ ms. It takes 18 floating point operations/0.25 ms, hence 72 operations/1 ms of simulation. Since the model is a 2-D system of ODEs without a reset, it cannot exhibit autonomous chaotic dynamics or bursting. Adding noise to this, or some other 2-D models, allows for stochastic bursting.

H. Hindmarsh–Rose

The Hindmarsh–Rose model of thalamic neuron [23] can be written

$$\begin{aligned} v' &= u - F(v) + I - w \\ u' &= G(v) - u \\ w' &= \frac{(H(v) - w)}{\tau} \end{aligned}$$

where F , G , and H are some functions. Depending on their choice, the model can, in principle, exhibit all of the neuro-computational properties in Fig. 1. The problem is, of course, how to find the functions to model, say RS or LTS neurons. Let us assume that this problem is somehow solved and that the functions are polynomials of the third degree (in the best case). Since we need to simulate the shape of the action potential, the maximal time step is 0.25 ms. Since it takes 30 floating point operations/0.25 ms of simulation time, it would take 120 operations to simulate 1 ms of the model. Again, this is an optimistic assessment that might never be achieved.

I. Morris–Lecar

Morris and Lecar [21] suggested a simple 2-D model to describe oscillations in barnacle giant muscle fiber. Because it has biophysically meaningful and measurable parameters, the model became quite popular in computational neuroscience community. It consists of a membrane potential equation with instantaneous activation of Ca current and an additional equation describing slower activation of K current

$$\begin{aligned} C\dot{V} &= I - g_L(V - V_L) - g_{Ca}m_\infty(V) \\ &\quad \times (V - V_{Ca}) - g_K n(V - V_K) \\ \dot{n} &= \lambda(V)(n_\infty(V) - n) \end{aligned}$$

where

$$\begin{aligned} m_\infty(V) &= \frac{1}{2} \left\{ 1 + \tanh \left[\frac{(V - V_1)}{V_2} \right] \right\} \\ n_\infty(V) &= \frac{1}{2} \left\{ 1 + \tanh \left[\frac{(V - V_3)}{V_4} \right] \right\} \\ \lambda(V) &= \bar{\lambda} \cosh \left[\frac{(V - V_3)}{(2V_4)} \right] \end{aligned}$$

with parameters: $C = 20 \mu\text{F}/\text{cm}^2$, $g_L = 2 \text{ mmho}/\text{cm}^2$, $V_L = -50 \text{ mV}$, $g_{Ca} = 4 \text{ mmho}/\text{cm}^2$, $V_{Ca} = 100 \text{ mV}$, $g_K = 8 \text{ mmho}/\text{cm}^2$, $V_K = -70 \text{ mV}$, $V_1 = 0 \text{ mV}$, $V_2 = 15 \text{ mV}$, $V_3 = 10 \text{ mV}$, $V_4 = 10 \text{ mV}$, $\bar{\lambda} = 0.1 \text{ s}^{-1}$, and applied current $I(\mu\text{A}/\text{cm}^2)$.

The model can exhibit various types of spiking, but could exhibit tonic bursting only when an additional equation is added, e.g., slow inactivation of Ca current. In this case, the model becomes equivalent to the Hodgkin–Huxley model discussed below (both have transient inward and persistent outward currents).

Because one needs to simulate the shape of the action potential in the Morris–Lecar model, the time step must be significantly smaller than 1 ms. We found that $\tau = 0.1$ ms is the largest step that gives reasonable results when the model is used to simulate cortical spiking neurons. Since the model involves hyperbolic tangents and exponents, it takes around 60 floating point operations to evaluate one 0.1-ms time step, which leads to 600 operations/1 ms of simulation.

J. Wilson Polynomial Neurons

Wilson [25] suggested to model cortical neurons using polynomial equations. His model consists of four differential equations, which we do not provide here. It can exhibit all neuro-computational properties in Fig. 1, provided that the parameters are chosen appropriately, which is not an easy task. The suggested time step in the model was 0.1 ms, though it could be pushed up to 0.25 ms without significant loss of precision or noticeable distortion of the shape of the action potential. It takes 45 floating point operations to evaluate 0.25 ms of the model, hence 180 operations/1 ms.

K. Hodgkin–Huxley

The Hodgkin–Huxley model [9] is one of the most important models in computational neuroscience. It consists of four equations and tens of parameters, not provided here, describing membrane potential, activation of Na and K currents, and inactivation of Na current. Though the model has quite limited behavior for original values of parameters, it can actually exhibit all properties in Fig. 1, if the parameters are tuned.

In general, scientists refer to all conductance-based models as being of the Hodgkin–Huxley-type. Such models are important not only because their parameters are biophysically meaningful and measurable, but also because they allow us to investigate questions related to synaptic integration, dendritic cable filtering, effects of dendritic morphology, the interplay between ionic currents, and other issues related to single cell dynamics.

The model is extremely expensive to implement. It takes 120 floating point operations to evaluate 0.1 ms of model time (assuming that each exponent takes only ten operations), hence, 1200 operations/1 ms. Thus, one can use the Hodgkin–Huxley formalism only to simulate a small number of neurons or when simulation time is not an issue.

IV. CONCLUSION

As the reader can see in Fig. 2, many models of spiking neurons have been proposed. Which one to choose? The answer depends on the type of the problem. If the goal is to study how the neuronal behavior depends on measurable physiological parameters, such as the maximal conductances, steady-state (in)activation functions and time constants, then the Hodgkin–Huxley-type model is the best. Of course, you could simulate only tens of coupled spiking neurons in real time.

In contrast, if you want to simulate thousands of spiking neurons in real time with 1 ms resolution, then there are plenty of models to choose from. The most efficient is the I&F model. However, the model cannot exhibit even the most fundamental properties of cortical spiking neurons, and for this reason it should be avoided by all means. The only advantage of the I&F model is that it is linear, and hence amenable to mathematical analysis. If no attempts to derive analytical results are made, then there is no excuse for using this model in simulations.

The quadratic I&F model is practically as efficient as the linear one, and it exhibits many important properties of real neurons, such as spikes with latencies, and bistability of resting and tonic spiking modes. However, it is 1-D, and hence, it cannot burst and cannot exhibit spike frequency adaptation. Thus, it can be used in simulations of cortical neural networks only when biological plausibility is not a great concern.

If the goal is to understand the fine temporal structure of cortical spike trains, and to use spike-timing as an additional variable to understand how the mammalian neocortex processes information, a spiking model that can exhibit all or most of the 20 neuro-computational properties of biological neurons summarized in Fig. 1 is required. The model recently proposed by Izhikevich [15] was developed exactly for these purposes. It is the simplest possible model that can exhibit all the firing patterns in Fig. 1. Indeed, removal of the (2) makes it 1-D with no possibility for bursting, removal of the term v^2 makes it linear and equivalent to the resonate-and-fire model.

The author has used the model to simulate a fully connected network of 1,000 cortical spiking neurons in real time with 1 ms resolution using only modest computational resources (1 GHz desktop PC; MATLAB code is provided in [15]. Using the C programming language, it is possible to speed up simulations by the factor of twenty). The network exhibited rhythms in the alpha and gamma frequency range, transient and sustained spike synchrony, spindle waves, sleep oscillations, and other temporal phenomena.

This model was also used in a simulation of a network of 100 000 spiking neurons with realistic cortical anatomy, axonal delays, and spike-timing dependent synaptic plasticity (STDP) [13]. Due to the interplay between spiking, plasticity, and delays, the neurons self-organized into *polychronous*

(i.e., multiple-timing) groups that could generate persistent time-locked firing patterns with ms precision. There are many such polychronous groups coexisting at the same time. In another study [12], we found that each neuron participates in many polychronous groups, so that the total number of groups the model could memorize can significantly exceed the number of neurons, or even the number of synapses in the network, resulting in unprecedented memory capacity.

In conclusion, having a network of computationally efficient and biologically plausible cortical spiking neurons interconnected according to the principles of known anatomy of the neocortex should be the goal of every scientist exploring information processing in the mammalian brain.

ACKNOWLEDGMENT

The author would like to thank N. S. Desai, A. K. Seth, J. R. Iversen, Y. Chen, J. L. McKinstry, E. C. Walcott, R. Muresan, R. Kavasseri, and M. Stich who read the first draft of the manuscript and made a number of useful suggestions.

REFERENCES

- [1] B. W. Connors and M. J. Gutnick, "Intrinsic firing patterns of diverse neocortical neurons," *Trends Neurosci.*, vol. 13, pp. 99–104, 1990.
- [2] G. B. Ermentrout, "Type I membranes, phase resetting curves, and synchrony," *Neural Comput.*, vol. 8, pp. 979–1001, 1996.
- [3] G. B. Ermentrout and N. Kopell, "Parabolic bursting in an excitable system coupled with a slow oscillation," *SIAM J. Appl. Math.*, vol. 46, pp. 233–253, 1986.
- [4] R. FitzHugh, "Impulses and physiological states in models of nerve membrane," *Biophys. J.*, vol. 1, pp. 445–466, 1961.
- [5] W. Gerstner and W. M. Kistler, *Spiking Neuron Models*. Cambridge, U.K.: Cambridge Univ. Press, 2002.
- [6] J. R. Gibson, M. Belerlein, and B. W. Connors, "Two networks of electrically coupled inhibitory neurons in neocortex," *Nature*, vol. 402, pp. 75–79, 1999.
- [7] C. M. Gray and D. A. McCormick, "Chattering cells: Superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex," *Science*, vol. 274, no. 5284, pp. 109–113, 1996.
- [8] A. L. Hodgkin, "The local electric changes associated with repetitive action in a nonmedulated axon," *J. Physiol.*, vol. 107, pp. 165–181, 1948.
- [9] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and application to conduction and excitation in nerve," *J. Physiol.*, vol. 117, pp. 500–544, 1954.
- [10] F. C. Hoppensteadt and E. M. Izhikevich, *Weakly Connected Neural Networks*. New York: Springer-Verlag, 1997.
- [11] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*, to be published.
- [12] E. M. Izhikevich, "Simple model of spiking network," *IEEE Trans. Neural Networks*, to be published.
- [13] E. M. Izhikevich, J. A. Gally, and G. M. Edelman, "Spike-timing dynamics of neuronal groups," *Cerebral Cortex*, vol. 14, pp. 933–944, 2004.
- [14] E. M. Izhikevich, N. S. Desai, E. C. Walcott, and F. C. Hoppensteadt, "Bursts as a unit of neural information: Selective communication via resonance," *Trends Neurosci.*, vol. 26, pp. 161–167, 2003.
- [15] E. M. Izhikevich, "Simple model of spiking neurons," *IEEE Trans. Neural Networks*, vol. 14, pp. 1569–1572, Nov. 2003.
- [16] —, "Resonate-and-fire neurons," *Neural Netw.*, vol. 14, pp. 883–894, 2001.
- [17] —, "Neural excitability, spiking, and bursting," *Int. J. Bifurcation Chaos*, vol. 10, pp. 1171–1266, 2000.
- [18] —, "Class 1 neural excitability, conventional synapses, weakly connected networks, and mathematical foundations of pulse-coupled models," *IEEE Trans. Neural Networks*, vol. 10, pp. 499–507, May 1999.
- [19] P. E. Latham, B. J. Richmond, P. G. Nelson, and S. Nirenberg, "Intrinsic dynamics in neuronal networks. I. Theory," *J. Neurophysiol.*, vol. 83, pp. 808–827, 2000.

- [20] J. Lisman, "Bursts as a unit of neural information: Making unreliable synapses reliable," *Trends Neurosci.*, vol. 20, pp. 38–43, 1997.
- [21] C. Morris and H. Lecar, "Voltage oscillations in the barnacle giant muscle fiber," *Biophys. J.*, vol. 35, pp. 193–213, 1981.
- [22] J. Rinzel and G. B. Ermentrout, "Analysis of neural excitability and oscillations," in *Methods in Neuronal Modeling*, C. Koch and I. Segev, Eds. Cambridge, MA: MIT Press, 1989.
- [23] R. M. Rose and J. L. Hindmarsh, "The assembly of ionic currents in a thalamic neuron. I The three-dimensional model," *Proc. R. Soc. Lond. B*, vol. 237, pp. 267–288, 1989.
- [24] G. D. Smith, C. L. Cox, S. M. Sherman, and J. Rinzel, "Fourier analysis of sinusoidally driven thalamocortical relay neurons and a minimal integrate-and-fire-or-burst model," *J. Neurophysiol.*, vol. 83, pp. 588–610, 2000.
- [25] H. R. Wilson, "Simplified dynamics of human and mammalian neocortical neurons," *J. Theor. Biol.*, vol. 200, pp. 375–388, 1999.



Eugene M. Izhikevich was born in Moscow, Russia, in 1967. He received the Master's degree in applied mathematics and computer sciences from Lomonosov Moscow State University in 1992, and the Ph.D. degree in mathematics from Michigan State University, East Lansing, in 1996.

Eugene is an Associate Fellow at The Neurosciences Institute, La Jolla, CA, where he studies nonlinear dynamics of biological neurons and builds large-scale models of the brain.

Theory of orientation tuning in visual cortex

(neural networks/cross-correlations/symmetry breaking)

R. BEN-YISHAI*, R. LEV BAR-OR*, AND H. SOMPOLINSKY†

*Racah Institute of Physics and Center for Neural Computation, Hebrew University, Jerusalem 91904, Israel; and †AT&T Bell Laboratories, Murray Hill, NJ 07974

Communicated by Pierre C. Hohenberg, AT&T Bell Laboratories, Murray Hill, NJ, December 21, 1994 (received for review July 28, 1994)

ABSTRACT The role of intrinsic cortical connections in processing sensory input and in generating behavioral output is poorly understood. We have examined this issue in the context of the tuning of neuronal responses in cortex to the orientation of a visual stimulus. We analytically study a simple network model that incorporates both orientation-selective input from the lateral geniculate nucleus and orientation-specific cortical interactions. Depending on the model parameters, the network exhibits orientation selectivity that originates from within the cortex, by a symmetry-breaking mechanism. In this case, the width of the orientation tuning can be sharp even if the lateral geniculate nucleus inputs are only weakly anisotropic. By using our model, several experimental consequences of this cortical mechanism of orientation tuning are derived. The tuning width is relatively independent of the contrast and angular anisotropy of the visual stimulus. The transient population response to changing of the stimulus orientation exhibits a slow “virtual rotation.” Neuronal cross-correlations exhibit long time tails, the sign of which depends on the preferred orientations of the cells and the stimulus orientation.

Neurons in the primary visual cortex respond preferentially to edges with a particular orientation. The input to the cortex is provided by neurons in the lateral geniculate nucleus (LGN), which respond independently of the stimulus orientation. The mechanism for the generation of orientation selectivity in the cortex is not fully known (1–8). According to the classical model of Hubel and Wiesel (1), the preferred orientation (PO) of a cortical cell originates from the geometrical alignment of the circular receptive fields of the LGN neurons that are afferent to it. The experimental evidence of this model is ambiguous. The alignment of the receptive fields of the LGN inputs to a cortical cell apparently parallels the cell's PO (2). However, suppression of cortical inhibition tends to considerably degrade orientation tuning (3, 4), suggesting that cortical circuitry plays an important role in shaping the relatively sharp orientation tuning in the cortex. Furthermore, estimates based on intracellular measurements indicate that most of the orientation-selective excitatory input to cortical cells comes from cortical feedback (6). Several models implicating cortical inhibition have been proposed (for reviews, see refs. 7 and 8). Their analysis, however, is hindered by the need to resort to massive numerical simulations of the different models. Computational complexity also precluded a detailed study of the role of the cortical excitatory connections. An important constraint on modeling orientation selectivity, which has not yet been fully addressed, is the experimental finding that the tuning width is relatively insensitive to the contrast of the stimulus (9). Most experimental and theoretical studies of orientation selectivity focus on the response properties of single neurons. However, valuable insight into the cooperat-

ivity among cortical neurons can be gained from measurements of the correlations between the responses of different neurons (10). Theoretical predictions regarding the magnitude and form of correlation functions in neuronal networks have been lacking.

Here we study mechanisms for orientation selectivity by using a simple neural network model that captures the gross architecture of primary visual cortex. By assuming simplified neuronal stochastic dynamics, the network properties have been solved analytically, thereby providing a useful framework for the study of the roles of the input and the intrinsic connections in the formation of orientation tuning in the cortex. Furthermore, by using a recently developed theory of neuronal correlation functions in large stochastic networks, we have calculated the cross-correlations (CCs) between the neurons in the network. We show that different models of orientation selectivity may give rise to qualitatively different spatiotemporal patterns of neuronal correlations. These predictions can be tested experimentally.

Model

We consider a network model with an architecture of a cortical hypercolumn. It consists of N_E excitatory neurons and N_I inhibitory neurons that respond selectively to a small oriented visual stimulus in a common receptive field. The neurons are parameterized by an angle θ , ranging from $-\pi/2$ to $+\pi/2$, that denotes their POs. The interaction between a presynaptic excitatory neuron θ' and a postsynaptic excitatory or inhibitory neuron θ is equal to $N_E^{-1}E(\theta - \theta')$. Similarly, the interaction between a presynaptic inhibitory neuron and a postsynaptic neuron is $N_I^{-1}I(\theta - \theta')$. The functions $E(\theta - \theta')$ and $I(\theta - \theta')$ represent the fact that the strength of the interactions between two orientation columns depends on the difference between their POs. The synaptic input from the LGN to a neuron θ is denoted as $h^{\text{ext}}(\theta - \theta_0)$, where θ_0 is the orientation of the external stimulus. Neurons switch stochastically between a quiescent state [denoted as $S(\theta) = 0$], representing a state with background firing rate (≈ 5 spikes per sec), and an active state [$S(\theta) = 1$], representing a state with saturated firing rate (≈ 500 spikes per sec). The transition rates between these states are governed by a sigmoidal gain function, $g[h(\theta, t)]$, where $h(\theta, t)$ is the total synaptic input to the neuron θ at time t . By assuming that the number of neurons is large and that they cover uniformly all the angles, the activity profile in the network at time t can be represented by a continuous function $m(\theta, t)$, $0 \leq m \leq 1$. This function denotes the mean activity level, relative to the saturation activity level, of the neurons with POs in the neighborhood of θ , at time t . Note that since the synaptic inputs to the excitatory and the inhibitory neurons in our model are the same, $m(\theta, t)$ describes the mean activity levels of both types of neurons.

In the limit of a large network, $m(\theta, t)$ obeys the following deterministic mean-field equations:

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LGN, lateral geniculate nucleus; PO, preferred orientation; CC, cross-correlation.

$$\tau_0 \frac{d}{dt} m(\theta, t) = -m(\theta, t) + g[h(\theta, t)], \quad [1]$$

where τ_0 is a microscopic characteristic time (a few milliseconds), and the total synaptic input $h(\theta, t)$ is

$$h(\theta, t) = \int_{-\pi/2}^{+\pi/2} \frac{d\theta'}{\pi} J(\theta - \theta') m(\theta', t) + h^{\text{ext}}(\theta - \theta_0). \quad [2]$$

Here $J(\theta - \theta')$ represents the net interaction between neurons θ and θ' , i.e., $J(\theta - \theta') = E(\theta - \theta') + I(\theta - \theta')$. We choose $E(\theta) = E_0 + E_2 \cos(2\theta)$ and $I(\theta) = -I_0 - I_2 \cos(2\theta)$ with $E_0 \geq E_2 \geq 0, I_0 \geq I_2 \geq 0$. This choice implies that both the excitatory interactions and the inhibitory interactions are maximal for orientation columns with similar POs. This is in accord with the observation that the excitatory and inhibitory postsynaptic potentials to a cell have similar tuning (11, 12). The net interaction is

$$J(\theta) = -J_0 + J_2 \cos(2\theta), \quad [3]$$

with $J_0 = I_0 - E_0$ and $J_2 = E_2 - I_2$, both assumed to be positive. The constant J_0 represents a uniform all-to-all inhibition; J_2 measures the amplitude of the orientation-specific part of the interaction. Neurons with similar POs are more strongly coupled excitatorily than neurons with dissimilar ones. The external input is taken as

$$h^{\text{ext}}(\theta) = c[1 - \varepsilon + \varepsilon \cos(2\theta)], \quad [4]$$

with $0 \leq \varepsilon \leq 1/2$. The parameter ε denotes the magnitude of the angular anisotropy of the input. The case $\varepsilon = 0$ denotes the limit where the input to the cortex is fully isotropic. The coefficient c measures the average amplitude of the input and will be referred to as the stimulus contrast. Finally, we use a semilinear gain function, i.e., $g(h) = 0$ for $h \leq T$, $g(h) = \beta(h - T)$ for $h \geq T$, and $g(h) = 1$ for $h \geq T + \beta^{-1}$. The parameter T is the threshold, which will be taken as 1, and β is the gain factor. This reflects the experimental finding that neuronal responses exhibit a sharp threshold contrast below which they vanish. Above threshold the mean firing rate of the neuron increases roughly linearly with the contrast. The slope of this rise is represented in our model by the constant β . Finally, the response saturates at high contrasts (13). The particular functional forms of the interactions, the LGN input and the gain function were chosen for the sake of simplicity. More general forms yield a behavior that is qualitatively similar to the present model.

Mechanisms for Orientation Selectivity

Mean-Field Solution. In the steady state, $m(\theta)$ is of the form $m(\theta) = M(\theta - \theta_0)$, where $M(\theta)$ peaks at $\theta = 0$. It represents the activity profile of the network relative to the orientation of the stimulus. Thus, $M(\theta)$ also represents the tuning curve of a neuron, centered at its PO. This function is given by the solution of the mean-field equation $M(\theta) = g[H(\theta)]$ with $H(\theta)$ being the local field, Eq. 2, centered at θ_0 . Its value is $H(\theta) = (\varepsilon c + J_2 m_2) \cos(2\theta) + c(1 - \varepsilon) - J_0 m_0$, where m_0 and m_2 are the zeroth and the second Fourier coefficients of $M(\theta)$. The order parameter m_0 measures the neuronal activity averaged over the whole network, whereas m_2 measures the variation of this activity across the different columns. Self-consistent equations for m_0 and m_2 are derived by evaluating the Fourier transform of $M(\theta)$. The resultant form of $M(\theta)$ is $M(\theta) = \beta(\varepsilon c + J_2 m_2)[\cos(2\theta) - \cos(2\theta_c)]$ for $|\theta| < \theta_c$ and zero otherwise (see examples in Fig. 2). The cutoff angle θ_c , which is determined by the condition $H(\pm\theta_c) = 1$, is half of the full

width of the tuning curve and will be denoted as the tuning width. The above solution holds for suprathreshold contrast, $c > 1$. Otherwise, $M(\theta)$ is zero. In the following, we study the orientation tuning properties of the steady-state solution. In particular, we examine the dependence of the tuning width on the stimulus contrast, c , and anisotropy, ε , for different ranges of values of the connectivity parameters.

Hubel and Wiesel Scenario. The simplest scenario for orientation tuning is based on the Hubel–Wiesel mechanism. In the context of our model, this scenario corresponds to the case where the only synaptic input is coming from the LGN and this input is strongly anisotropic. Hence, the cortical local interaction $J(\theta)$ is zero and ε is substantially bigger than zero. The dependence of the tuning width θ_c on c in this case is shown in Fig. 1A for different values of ε . Near threshold (i.e., for $c \rightarrow 1$), θ_c vanishes, signaling the sharpening of the tuning width, relative to the input, by the thresholding operation of the cortical neurons. This, however, results in a strong dependence of the width on the contrast. As c increases, θ_c grows and beyond a certain value saturates at its maximum value $\pi/2$, which corresponds to the situation where all the neurons are activated by all stimuli.

Uniform Inhibition. A simple cortical mechanism for sharpening the orientation tuning is based on isotropic cortical inhibition. This scenario corresponds in our model to $J_0 > 0$ and $J_2 = 0$. In this case, the external input of each neuron has to overcome an effective enhanced threshold that increases with c because of the increase of the inhibitory feedback. Thus, even for $c \gg 1$, the inhibition may provide a sufficiently potent threshold to sharpen the tuning width. This is clearly seen in Fig. 1B (dashed line). Nevertheless, there is a strong dependence of the width on the contrast c . In addition, the width depends strongly on the anisotropy of the input, ε . In particular, when ε is small, $\theta_c = \pi/2$ for large c . This is shown in Fig. 1C (dashed line).

Marginal Phase. In the cases studied above, the anisotropy of the input to the cortex is the only source of orientation specificity. Can orientation specificity be generated by cortical interactions alone—even in the absence of significant anisotropy in the external input? To answer this question, we have studied the case of $\varepsilon = 0$ with J_0 and $J_2 > 0$. The assumed rotational symmetry of the cortex implies that when $\varepsilon = 0$, there is always a homogeneous solution, $M(\theta) = m_0$, which agrees with the naive expectation that the orientation tuning disappears when the external stimulus is isotropic. However, the equations yield also an inhomogeneous solution, provided that the angular modulation of the cortical interactions, J_2 , is big enough, $J_2 > 2/\beta$. This solution represents a spontaneous generation of orientation tuning. It can be written as $m(\theta) = M(\theta - \phi)$, where the angle ϕ , which determines the location of the peak in $m(\theta)$, is arbitrary. This means that there is a continuum of stable states: they all have identical orientation tuned activity profiles, but they differ in the location of their peaks. The overall magnitude of $m(\theta)$ is determined by c but the shape of the activity profile is determined only by the cortical interactions. In particular, the width of $m(\theta)$ is independent of c but strongly depends on the magnitude of J_2 . The bigger J_2 , the smaller the width. We denote this solution a marginal phase because there are no barriers between the different attractors. This implies that the motion from one attractor to another is easier than motions in other directions of phase space. When the marginal solution exists, the naive homogeneous solution is unstable.

In reality, the location of the peak of the activity is not arbitrary but is determined by the orientation of the external input. Therefore, in a realistic implementation of the last scenario ε should be assumed to be small but nonzero. Since ε is small, the main effect of the anisotropy of the input is to select among the continuum of possible states that state in which the location of the peak in the activity matches the

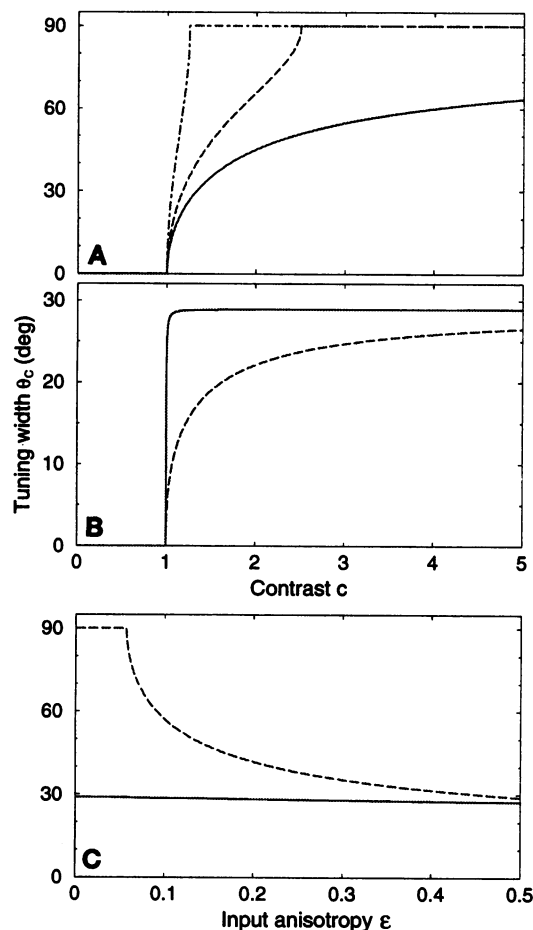


FIG. 1. Dependence of the orientation tuning with θ_C on the stimulus contrast and input anisotropy, for different mechanisms of tuning. The angle θ_C is the half-width of the tuning curve: neurons with POs of $>\theta_C$ away from the stimulus orientation are not activated. (A) The Hubel-Wiesel model, where the cortical interaction parameters J_0 and J_2 are zero. The tuning width is plotted against the contrast of the visual stimulus, c , for different values of the input anisotropy parameter, ϵ : $\epsilon = 0.1, 0.3$, and 0.5 (dot-dashed, dashed, and solid lines, respectively). For $\epsilon < 0.5$, there is a critical value of c above which $\theta_C = \pi/2$, meaning that all neurons are activated by all stimuli. In the range where θ_C is small, it strongly depends on the level of contrast. Note that $\epsilon = 0.5$ is the maximal value of anisotropy in our model. Here and in Figs. 2 and 3, $\beta = 0.1$. (B) Effect of cortical interactions on the tuning width. Dashed curve is for uniform cortical inhibition, with $\epsilon = 0.5$, $J_0 = 155$, and $J_2 = 0$. The solid line is for parameters close to the marginal phase, where the tuning is dominated by the cortical interactions. In this case, θ_C saturates quickly to $\approx 30^\circ$, while the level of maximal activity is still far from saturation (data not shown). Parameters are $\epsilon = 0.01$, $J_0 = 86$, and $J_2 = 112$. They were chosen so that the value of θ_C agrees with the average value of the half-width for simple cells in the cat visual cortex (5). (C) The tuning width as a function of the input anisotropy in the limit $c \gg 1$. The solid line is for the marginal phase; the dashed curve is for the Hubel-Wiesel mechanism with uniform cortical inhibition. Parameters are as in B.

orientation of the stimulus, but it will not affect the shape of the tuning curve very much. This is indeed borne out by a solution of the mean-field equations with a small nonzero ϵ and large J_0 and J_2 . There is a unique stable state, $m(\theta) = M(\theta - \theta_0)$, where $M(\theta)$ is approximately the same as for $\epsilon = 0$. Whereas when $\epsilon = 0$, θ_C was independent of c , and when ϵ is nonzero, θ_C vanishes as $c \rightarrow 1$. However, this decrease in θ_C occurs in a very narrow range of c near threshold, $c - 1 = O(\epsilon)$, as shown in Fig. 1B (solid line). The value of θ_C for larger c is independent of c . It is also insensitive to changes in the value of ϵ , as shown in Fig. 1C (solid line).

Virtual Rotation

An interesting difference between the scenarios described above appears in the time-dependent response of the system to a change in the orientation of the external stimulus. Consider the case where the orientation of the external stimulus is changed at time 0 from $\theta_0 = \theta_1$ to $\theta_0 = \theta_2$. We assume that initially the stimulus was present long enough so that the system reached a steady-state $m(\theta)$ with a peak at $\theta = \theta_1$. How will the population activity respond to the change in orientation? In the case of isotropic cortical interactions ($J_2 = 0$), the initial activity profile that peaked at θ_1 decays in magnitude while another profile peaked at θ_2 grows, as shown in Fig. 2A. In the marginal phase, the response is drastically different. The population activity is $m(\theta, t) \approx M[\theta - \phi(t)]$, where $M(\theta)$ is the activity profile at steady state. We denote this solution as virtual rotation. The activity at time t is similar to that that

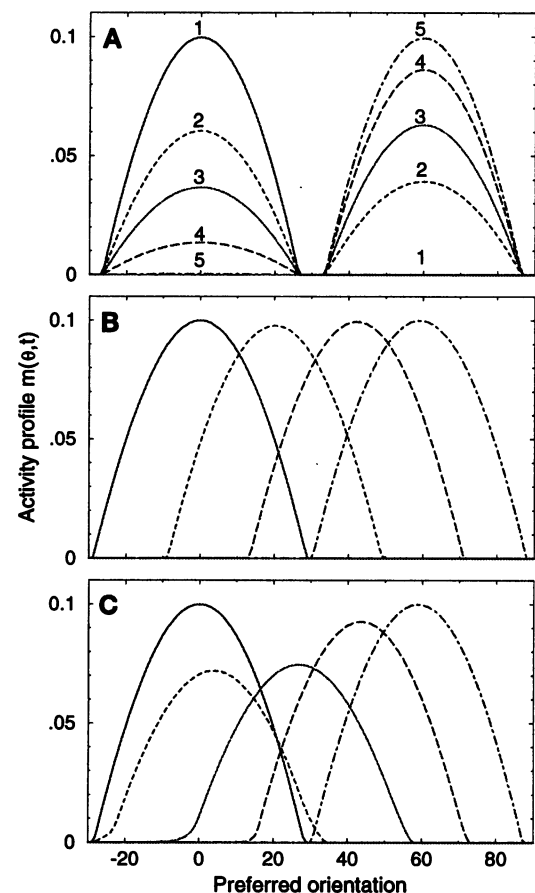


FIG. 2. Evolution of neuronal activity in response to a change in the stimulus orientation from an initial value $\theta_1 = 0^\circ$ to $\theta_2 = 60^\circ$. The change occurs at $t = 0$. (A) Hubel-Wiesel model with uniform inhibition. The activity profile centered at 0° decays while a new one, centered around 60° , grows. Neurons in intermediate columns (e.g., with $\theta = 30^\circ$) stay in the quiescent state during the whole process. The time constant of the decay and growth is τ_0 . Times are 0, 0.5, 1, 2, and 6 τ_0 (lines 1–5, respectively). Parameters are as in Fig. 1B (dashed line) and $c = 5$. For this value of contrast, the maximal activity level relative to saturation is 0.1, which corresponds to a rate of ≈ 50 spikes per sec. (B) Virtual rotation in the marginal phase. The activity profile moves toward θ_2 , activating successively the intermediate columns, and undergoing only very small changes during the process. $c = 1.678$; other parameters are as in Fig. 1B (solid line). Times are (left to right) 0, 90, 210, and 600 τ_0 . (C) Virtual rotation accompanied by deformations in the activity profile. The magnitude of the deformations depends on the magnitude of ϵ . Here $\epsilon = 0.1$, $J_0 = 73$, $J_2 = 110$, and $c = 1.45$. Times are (left to right) 0, 2, 10, 20, and 60 τ_0 . By assuming $\tau_0 \approx 5$ msec, the predicted rotation velocity is 200 to 400°/sec.

would occur if there was an external stimulus with an orientation $\theta_0 = \phi(t)$. Substituting this form for $m(\theta, t)$ in the dynamic equations (Eqs. 1 and 2), we find that the variable $\phi(t)$ obeys $\dot{\phi}(t) \approx -\omega_0 \sin[2\phi(t) - 2\theta_2]$, with $\phi(t = 0) = \theta_1$. The angular frequency ω_0 is given by

$$\omega_0 \tau_0 = \varepsilon c / (2J_2 m_2), \quad [5]$$

where m_2 is the second Fourier coefficient of $M(\theta)$. Note that in the marginal phase m_2 remains finite even for small ε . Thus, ω_0 , the angular velocity of the activity profile, is determined by the ratio of the modulation amplitudes of the external input and of the cortical synaptic input. This result is valid provided $\omega_0 \tau_0 = O(\varepsilon) \ll 1$. An example is shown in Fig. 2B, which corresponds to a marginal phase with $\varepsilon = 0.01$. The activity profile moves with an (initial) angular velocity $0.2^\circ/\tau_0$, without significant change in its shape. Fig. 2C shows the case of $\varepsilon = 0.1$. Here, the angular velocity is faster by an order of magnitude, as predicted by Eq. 5. In this case, the propagation of the activity is accompanied by significant transient changes in the amplitude of the activity profile.

CCs

The CC function $C(\theta, \theta'; \tau)$ measures the correlation between the temporal fluctuations in the rate of activity of a neuron θ at time t and that of another neuron θ' , at $t + \tau$. In our model, these fluctuations arise from the underlying stochasticity of the neuronal dynamics. The deterministic mean-field equations, Eq. 1, neglect these small fluctuations. To study the CCs induced by these fluctuations, we have applied a recently developed theory of correlations in stochastic neural networks (14) and calculated analytically the CCs in our model for various parameter regimes. We consider here the CCs between excitatory neurons in the case of uniform inhibition and in the marginal phase. In the uniform inhibition case, the CCs are independent of θ , θ' , and θ_0 , as long as the two neurons are activated by the stimulus, and the CCs decay on a fast time scale, of the order of the microscopic time constant τ_0 . This is shown in Fig. 3A. The results for the marginal phase are drastically different. Here, in addition to fast-decaying components, the CCs possess also a slow component that has the following unique dependence on θ , θ' , and θ_0 ,

$$C(\theta, \theta', \tau) \propto \frac{1}{\varepsilon} \sin(2\theta - 2\theta_0) \sin(2\theta' - 2\theta_0) \exp(-\varepsilon|\tau|/\tau_0), \quad \tau \gg \tau_0. \quad [6]$$

Thus, the longest decay time of the CCs is τ_0/ε , where ε is the amplitude of the angular modulation of the input, which is assumed to be small in this phase. The magnitude of the slow component is big, on the order of ε^{-1} . This slow component results from the fact that in the marginal phase, the noise induces random slow wandering of the system between neighboring attractors, thus generating small random virtual rotations.

The result (Eq. 6) implies that not only the magnitude but also the sign of the CC between a pair of neurons may depend on the stimulus orientation. If the stimulus orientation is larger or smaller than both θ and θ' , the CC is positive. On the other hand, if θ_0 is intermediate between the two POs, the CC is negative, despite the fact that the direct interaction between the two neurons is positive. These results are shown in Fig. 3B–D, where we present the full CCs in the marginal phase. Comparing these results with Eq. 6, it is seen that the properties of the CCs at short τ are affected by the contributions from the fast-decaying noncritical modes of fluctuations, which generate a narrow positive peak near the center. The long-time component is dominated by the critical mode.

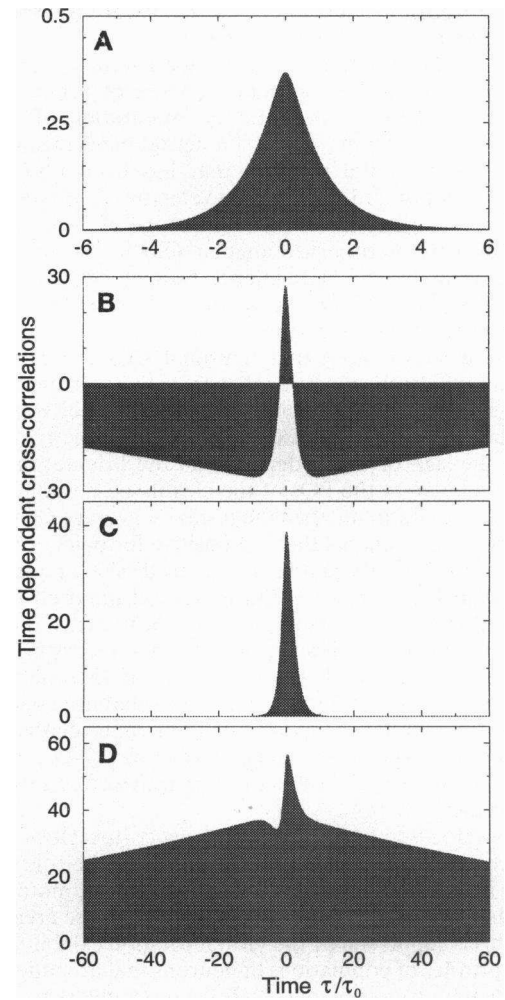


FIG. 3. Time-dependent CC between the fluctuations in the activity of two excitatory neurons, $\theta = -10^\circ$ and $\theta' = 10^\circ$, responding to a common oriented stimulus. (A) Uniform inhibition. The CC decay has a time constant on the order of τ_0 . The CCs are independent of the stimulus orientation θ_0 (data not shown), as long as both neurons are activated by it. Parameters are as in Fig. 2A. The vertical scale displays the CCs multiplied by the number of excitatory neurons in the network, N_E . (B–D) CCs in the marginal phase for different stimulus orientations. Parameters are as in Fig. 2B. The CCs exhibit a slow-decaying component, with a time constant, τ_0/ε , and a magnitude that scales as ε^{-1} . (B) $\theta_0 = 0^\circ$, i.e., between θ and θ' , which results in a negative slow component, as predicted by Eq. 6. (C) $\theta_0 = 10^\circ$, for which the slow component vanishes. (D) $\theta_0 = 16^\circ$, for which the slow component becomes positive.

Discussion

By using a simple neural network model, we have studied theoretically the consequences of different mechanisms for orientation selectivity in visual cortex. The network architecture is consistent with the known anatomy and physiology of visual cortex. In particular, the assumed angular modulation of the intracortical interactions, Eq. 3, is supported by the spatial distributions of the dendritic and axonal arborizations in primary visual cortex (8). We find that if the amplitude of the cortical angular modulation, J_2 , is sufficiently strong, the network can be in a state where the orientation tuning is dominated by the cortical circuitry. This state has several characteristic properties that can be verified experimentally. First, the tuning width is largely independent of stimulus properties such as its contrast and its degree of anisotropy. Indeed, invariance of the tuning width to contrast has been well documented (9). Measuring the effect of decreasing the

angular anisotropy of a stimulus on the orientation selectivity would be an additional important test.

The second characteristic, the virtual rotation, can be verified by measuring the transient response of primary visual cortex to an abrupt change in the orientation of a visual stimulus. Our model may suggest a neural mechanism for the psychophysical mental rotation that has been observed in object recognition (15). The initial velocities displayed in Fig. 2 *B* and *C*, assuming that τ_0 is in the range of 5–10 msec, are consistent with the observed angular velocity of mental rotation in psychophysical experiments, ranging between 60°/sec (15) and 400°/sec (16). Of course, other neuronal mechanisms for mental rotation are possible.

We have shown here that neuronal CCs are potentially important indicators of neuronal cooperativity. In our case, we predict that if cortical circuitry plays a dominant role in the orientation-selective response, CCs should exhibit a slow component, the sign of which depends on the orientation of the stimulus relative to the POs of the correlated neurons. If one averages over all stimuli, then the resultant average CCs should exhibit slow components that are positive for pairs of neurons with similar POs, and negative for largely dissimilar ones. Long time tails in CCs, which extend to several hundred milliseconds, are frequently observed in cortical areas, including primary visual cortex. Testing our theory requires systematic measurements of the slow components of CCs of pairs of neurons in different orientation columns that are coactivated by the same stimulus. Negative CCs, although relatively rare, have been observed in several cortical areas (17–20). Some of these correlations may be due to cooperative effects similar to those predicted here.

We have focused in this work on visual cortex. However, our approach can be applied also to the study of the role of local cortical interactions in other areas, primarily in motor areas, although the nature of the synaptic inputs to these areas is less clear. This is supported by the virtual rotation exhibited by the activity profiles of populations of neurons that are tuned to the direction of limb movement (17). Indeed, a network model that is similar to ours in some aspects was recently proposed for that system (21). It should be noted that the virtual rotation exhibited by our model is a direct outcome of the marginality of the network state and does not involve an active modulation of synapses, as in ref. 21. Also, the study of the neuronal correlations requires a more complex dynamics than the smooth deterministic model of ref. 21. As shown by our work, the correlations between a pair of neurons are not directly related to the value of the interactions between them. Instead, they are a consequence of the cooperative dynamical fluctuations in the network to which they belong.

Finally, we note that the marginal phase in the present theory is a result of the underlying angular symmetry of our model. It is thus similar to marginal phases that appear in physical systems at thermal equilibrium whenever a continuous symmetry is spontaneously broken. In particular, these systems exhibit “critical transverse correlations” that are similar to the slowly decaying CCs predicted here (22).

We thank M. Abeles, D. Kleinfeld, S. Seung, R. Shapley, and M. Tsodyks for most helpful discussions. This research is partially supported by the United States–Israel Binational Science Foundation.

- Hubel, D. H. & Wiesel, T. N. (1962) *J. Physiol. (London)* **160**, 106–154.
- Chapman, B., Zahs, K. R. & Stryker, M. P. (1991) *J. Neurosci.* **11**, 1347–1358.
- Tsumoto, T., Eckart, W. & Creutzfeldt, O. D. (1979) *Exp. Brain Res.* **34**, 351–363.
- Sillito, A. M., Kemp, J. A., Milson, J. A. & Berardi, N. (1980) *Brain Res.* **194**, 517–520.
- Orban, G. A. (1984) *Neuronal Operations in the Visual Cortex* (Springer, Berlin).
- Ahmed, B. A., Anderson, J. C., Douglas, R. J., Martin, K. A. C. & Nelson, J. C. (1994) *J. Comp. Neurol.* **341**, 39–49.
- Ferster, D. & Koch, D. (1987) *Trends Neurosci.* **10**, 487–492.
- Martin, K. A. C. (1988) *Q. J. Exp. Physiol.* **73**, 637–702.
- Skottun, B. C., Bradley, A., Sclar, G., Ohzawa, I. & Freeman, R. D. (1987) *J. Neurophysiol.* **57**, 773–786.
- Fetz, E., Toyama, K. & Smith, W. (1991) in *Cerebral Cortex*, eds. Peters, A. & Jones, G. (Plenum, New York), Vol. 9.
- Ferster, D. (1986) *J. Neurosci.* **6**, 1284–1301.
- Douglas, R. J., Martin, K. A. C. & Whitteridge, D. (1991) *J. Physiol. (London)* **440**, 659–696.
- Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. (1981) *Exp. Brain Res.* **41**, 414–419.
- Ginzburg, I. & Sompolinsky, H. (1994) *Phys. Rev. E* **50**, 3171–3191.
- Shepard, R. N. & Metzler, J. (1971) *Science* **171**, 701–703.
- Georgopoulos, A. P. & Massey, J. T. (1987) *Exp. Brain Res.* **65**, 361–370.
- Georgopoulos, A. P., Taira, M. & Lukashin, A. (1993) *Science* **260**, 47–52.
- Fetz, E. E. & Shupe, L. E. (1994) *Science* **263**, 1295–1296.
- Hata, Y., Tsumoto, T., Sato, H., Hagihara, K. & Tamura, H. (1988) *Nature (London)* **335**, 815–817.
- Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovov, H. & Aertsen, A. (1995) *Nature (London)* **373**, 515–518.
- Lukashin, A. V. & Georgopoulos, A. P. (1993) *Biol. Cyber.* **69**, 517–524.
- Forster, D. (1975) *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions* (Benjamin, Reading, MA).

The Neuroscientist

<http://nro.sagepub.com>

A Neural Circuit Basis for Spatial Working Memory

Christos Constantinidis and Xiao-Jing Wang

Neuroscientist 2004; 10; 553

DOI: 10.1177/1073858404268742

The online version of this article can be found at:
<http://nro.sagepub.com/cgi/content/abstract/10/6/553>

Published by:



<http://www.sagepublications.com>

Additional services and information for *The Neuroscientist* can be found at:

Email Alerts: <http://nro.sagepub.com/cgi/alerts>

Subscriptions: <http://nro.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations (this article cites 115 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
<http://nro.sagepub.com/cgi/content/refs/10/6/553>

A Neural Circuit Basis for Spatial Working Memory

CHRISTOS CONSTANTINIDIS and XIAO-JING WANG

The maintenance of a mental image in memory over a time scale of seconds is mediated by the persistent discharges of neurons in a distributed brain network. The representation of the spatial location of a remembered visual stimulus has been studied most extensively and provides the best-understood model of how mnemonic information is encoded in the brain. Neural correlates of spatial working memory are manifested in multiple brain areas, including the prefrontal and parietal association cortices. Spatial working memory ability is severely compromised in schizophrenia, a condition that has been linked to prefrontal cortical malfunction. Recent computational modeling work, in interplay with physiological studies of behaving monkeys, has begun to identify microcircuit properties and neural dynamics that are sufficient to generate memory-related persistent activity in a recurrent network of excitatory and inhibitory neurons during spatial working memory. This review summarizes recent results and discusses issues of current debate. It is argued that understanding collective neural dynamics in a recurrent microcircuit provides a key step in bridging the gap between network memory function and its underlying cellular mechanisms. Progress in this direction will shed fundamental insights into the neural basis of spatial working memory impairment associated with mental disorders. *NEUROSCIENTIST* 10(6):553–565, 2004. DOI: 10.1177/1073858404268742

KEY WORDS Prefrontal cortex, Saccade, Neuron, Network, Schizophrenia, Dopamine

The ability to hold information in memory over a time scale of seconds is a critical component of cognitive functions such as language, abstract thought, planning and most of the mental processes we associate with human intelligence. Working memory, as this critical function is known, can be thought of as a blackboard of the mind (Baddeley 1992). Because of its central role in cognitive function, working memory has attracted intense research scrutiny over several decades. Early, lesion studies revealed profound short-term memory deficits after selective ablation of the prefrontal cortex (Jacobsen 1936; Milner 1963; Goldman-Rakic 1987). Subsequent neurophysiological experiments in monkeys demonstrated that prefrontal neurons continue to discharge even after the offset of transient sensory stimuli that animals are required to remember (Fuster and Alexander 1971; Kubota and Niki 1971; Funahashi and others 1989). This persistent discharge is widely thought to be the neural correlate of working memory. The

advent of imaging techniques, PET and fMRI, verified persistent activation of the human brain during the maintenance of information in memory and confirmed the involvement of the prefrontal cortex in the process (Jonides and others 1993; Courtney and others 1997; Ungerleider and others 1998). Neural correlates of working memory have since been observed in other brain regions as well, and indeed, the prefrontal cortex is only part of a broader network of interconnected cortical and subcortical areas (Fig. 1). The differential roles and relative contributions of these brain structures in memory function continue to be investigated.

The representation of spatial information and particularly the position of a visual stimulus on the fronto-parallel plane has been a popular model for the investigation of working memory. Among its advantages is that space is represented in the brain in a well-understood, parametric fashion; the primary visual cortex is organized retinotopically, providing a clear correspondence between neuronal activity and stimulus position. In contrast to retinal location, the representation of other stimulus attributes (e.g., object identity) is encoded in still poorly understood dimensions. Visuospatial memory tasks (Fig. 1) provide the best-understood model of working memory, and much effort has been devoted to understanding the cortical mechanisms underlying their execution. This review summarizes recent progress in the field, with an emphasis on close interactions between experimental and computational approaches that have

From the Department of Neurobiology and Anatomy, Wake Forest University School of Medicine, Winston-Salem, North Carolina (CC), and the Center for Complex Systems, Brandeis University, Waltham, Massachusetts (X-JW).

Supported by grants from the McDonnell Foundation (CC), National Institute of Mental Health (MH062349), and Swartz Foundation (X-JW).

Address correspondence to: Xiao-Jing Wang, Center for Complex Systems, Brandeis University, Waltham, MA 02254 (e-mail: xjwang@brandeis.edu).

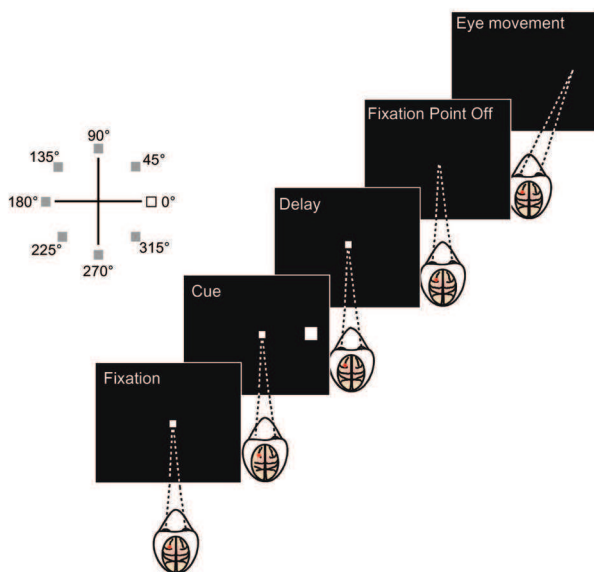


Fig. 1. Successive frames illustrate the sequence of events in the oculomotor delayed-response task. Trials begin with the appearance of a fixation point at the center of the screen, which the monkey is required to foveate throughout the trial. A spatial cue is subsequently presented, typically at one of eight locations (*left*). After a delay period of a few seconds, the fixation point is turned off and the monkey is required to indicate the location of the cue by moving his eyes accordingly on the screen.

led to new insights, as well as open questions for future investigations.

Neural Correlates of Spatial Working Memory

Electrophysiological studies in the monkey prefrontal cortex (Fig. 2) first revealed that a population of prefrontal neurons continues to discharge in a persistent fashion, even after sensory stimuli are no longer present (Fuster and Alexander 1971; Kubota and Niki 1971; Funahashi and others 1989). The part of the visual space where stimulus appearance can produce sustained activation has been termed the neuron's *memory field*, in analogy to the receptive field of neurons responding to sensory stimulation. Different prefrontal neurons exhibit different memory fields, and the overall activity of the population can encode the location of a remembered stimulus.

To elucidate the neural mechanisms of spatial working memory, it is desirable to dissect neural activity related to the mnemonic content per se, dissociated from the other aspects of a behavioral task. One of the most often-used, as well as simplest, behavioral paradigms for evaluating persistent activity is the oculomotor delayed-response task (Fig. 2). In this task, monkeys are required to fixate a central point on a screen, to remember where a brief visual cue appeared, and after a delay period of a few seconds, to indicate the remembered location of the cue by shifting their eyes to it. Neuronal activity persisting after the disappearance of the visual cue in such a task may, in principle, mediate the memory of the cue

location but may also be related solely to the preparation of the motor response. Indeed, one of the disadvantages of the visuospatial model is that spatial information in the brain is inexorably linked with motor circuits for the guidance of eye and limb movements to visual targets. It is a challenge, therefore, to dissociate experimentally observed neuronal activity related to visual spatial memory with that related to the preparation of motor commands. A number of experiments have sought to distinguish between the two alternatives by dissociating the location of a remembered stimulus with the direction of a motor response. Monkeys have been trained to perform an antisaccade task requiring an eye movement away from a remembered stimulus (Funahashi, Chafee, and others 1993), a conditional response task involving a movement in a direction other than the stimulus location (Niki and Watanabe 1976; Takeda and Funahashi 2002), or a spatial match to sample task, demanding a lever release when a stimulus appears at a previously cued location (Sawaguchi and Yamane 1999). In all cases, discharge of the majority of dorsolateral prefrontal neurons was found to encode the location of the remembered stimulus, although a minority of neurons did represent the actual motor response. A signal-detection theory analysis of prefrontal cortical activity revealed that the latter reflects sensory attributes of a remembered stimulus even when the particular stimulus is not guiding a motor action (Constantinidis and others 2001b).

Although visual stimuli have been most extensively used in neurophysiological investigation, spatially localized mnemonic neuronal activity is not restricted to vision. Prefrontal cortical neurons exhibit sustained discharges in auditory memory tasks, spatially tuned to the location of the sound source (Azuma and Suzuki 1984; Vaadia and others 1986; Bodner and others 1996). A particular prefrontal cortical subdivision appears to receive specialized inputs of spatial auditory information (Romanski and others 1999). Beyond the prefrontal cortex, neurons in lateral intraparietal area (LIP) of the posterior parietal cortex have been shown to be active during tasks requiring orienting to a remembered auditory target (Mazzoni and others 1996) but only if they have been trained to perform an eye-orienting task to auditory stimuli (Grunewald and others 1999).

Working Memory in a Distributed Network

Although first described in the prefrontal cortex, neurons active during spatial working memory have been reported in other cortical areas. The dorsolateral prefrontal cortex (areas 46 and 8) receives input from the posterior parietal cortex, an area of the dorsal visual stream, which is involved with the processing of visuospatial information. Posterior parietal neurons in areas LIP and 7a (Fig. 3) are tuned to the spatial location of stimuli, and they too discharge in a sustained fashion when animals are trained to remember the spatial location of a stimulus (Andersen and others 1987; Gnadt and Andersen 1988; Quintana and Fuster 1992; Constantinidis and Steinmetz 1996). A study comparing

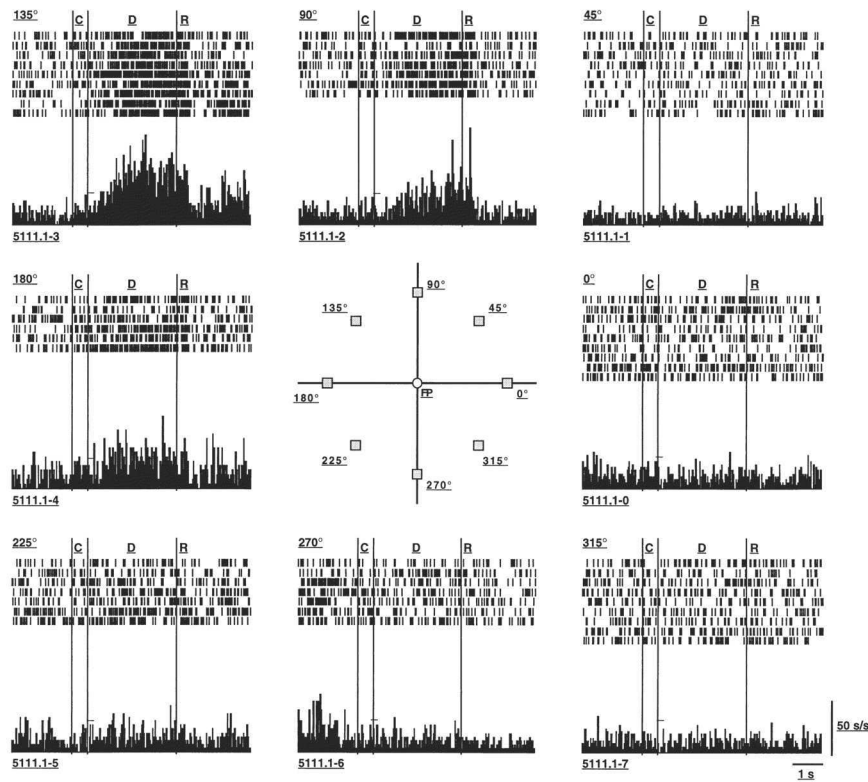


Fig. 2. Activity of a single prefrontal neuron, exemplifying persistent discharges during the execution of the oculomotor delayed-response task. Discharges are arranged as to indicate the location of the cue. The neuron is mostly active during the delay period following presentation of a stimulus in the upper left (135-degree) location. From Funahashi and others (1989) with permission.

responses in the dorsolateral prefrontal and posterior parietal cortex of the same monkeys trained to perform an oculomotor delayed-response task revealed no appreciable differences in the two areas (Chafee and Goldman-Rakic 1998). Indeed, imaging studies in humans almost invariably report concurrent prefrontal and parietal activation in working memory tasks (Jonides and others 1993; Courtney and others 1997; Owen and others 1998; Ungerleider and others 1998; Marshuetz and others 2000; Bunge and others 2001; Rowe and Passingham 2001; Stern and others 2001; Munk and others 2002). The distinct and cooperative roles of these areas remain unresolved.

Memory-related activity has been further described in earlier areas of the dorsal visual pathway projecting to the posterior parietal cortex, such as area V3A and MT (Nakamura and Colby 2000; Bisley and others 2004) and even V1 (Super and others 2001). In addition, spatial memory responses have been recorded in the primate entorhinal cortex (Suzuki and others 1997). Also active during spatial working memory are brain areas involved in the control of movement, such as the premotor and supplementary motor cortex (di Pellegrino and Wise 1993b; Russo and Bruce 1996), superior colliculus (Glimcher and Sparks 1992), and basal ganglia (Hikosaka and Wurtz, 1983; Hikosaka and others 1989; Kimura and others 1992). These results show that

mnemonic neural activity is widespread. However, information is scarce as to whether persistent activity is primarily generated in certain local brain areas, while neurons in other areas simply reflect sustained inputs from memory cells upstream, or persistent activity depends on feedback connection pathways in a large brain network. There have been conflicting reports on whether delay-period-persistent activity is prevalent in the thalamus (Fuster and Alexander 1973; Sommer and Wurtz 2004). The clarification of this issue would help to assess whether the cortico-thalamic-basal ganglia loop is critical to the generation of working memory activity in the cortex.

Important differences between the activation of the prefrontal and posterior parietal cortex have begun to be unveiled through the use of more complex cognitive tasks. When monkeys are trained to remember the spatial location of a sample stimulus and ignore intervening, distracting stimuli, neurons in posterior parietal areas 7a and LIP encode the location of the most recent stimulus, whether it is the remembered sample or the behaviorally irrelevant distractor. Parietal neurons respond in a persistent fashion to the presentation of distractor stimuli in the receptive field (Constantinidis and Steinmetz 1996; Powell and Goldberg 2000) while ceasing to represent the sample stimulus, although the animal continues to retain the sample in memory and successfully completes

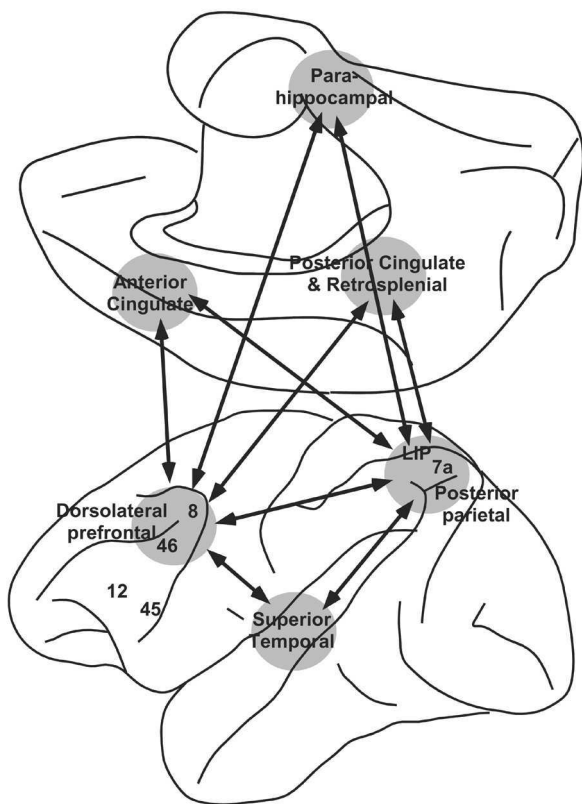


Fig. 3. The spatial working memory network. Cortical areas anatomically interconnected with the dorsolateral prefrontal cortex and activated by spatial memory.

the trial. Prefrontal neurons, on the other hand, encode the actively remembered cue (di Pellegrino and Wise 1993a, 1993b). This prefrontal trait of resistance to sequentially presented distractors has been recently verified by human imaging experiments (Cornette and others 2002; Sakai and others 2002). Computational studies have offered insights on how a working memory network can filter out behavioral irrelevant stimuli, suggesting a possible role of NMDA receptors, dopamine innervation, and specialized interneuron types (Compte and others 2000; Brunel and Wang 2001; Wang and others 2004).

Responses of neurons activated by spatial working memory can be modulated by other factors, as it has recently been revealed by the use of more complex behavioral paradigms. When the animal is cued to remember one of multiple stimuli in the visual field, prefrontal neurons preferentially represent the attributes of the attended stimulus (Rainer and others 1998b; Everling and others 2002). The response of prefrontal neurons to the same operant stimuli can also vary depending on the expectation of reward, when the later changes from trial to trial (Leon and Shadlen 1999). A similar modulation of neuronal responses to identical stimulation has been observed in animals trained to perform a number of alternative tasks, requiring association of the same stimulus with different possible motor

choices, according to different rules (White and Wise 1999; Asaad and others 2000; Wallis and others 2001). Prefrontal activation by such nonspatial factors has been readily observed in human fMRI experiments, leading some experimenters to question whether the role of prefrontal cortical activity is spatial in nature at all (Curtis and D'Esposito 2003). Such a view is difficult to reconcile, however, with the unequivocal spatial tuning of neuronal activity recorded in neurophysiological experiments. Effects of cognitive functions such as attention and reward expectation have been demonstrated in other areas of the working memory network, most extensively in the posterior parietal cortex (Platt and Glimcher 1999; Constantinidis and Steinmetz 2001). How spatial location can be decoded independently of other factors modulating neuronal activity is still an open question.

Microcircuit Organization of Spatial Working Memory

There has been some debate about whether the ventral subdivision of the prefrontal cortex comprising areas 12 and 45 (the prefrontal inferior convexity) might have a role in spatial memory as well. These areas receive input from the inferior temporal cortex, the final stage of the ventral visual pathway, and for this reason have been thought to represent memory for the identity of stimuli (Wilson and others 1993). However, it has been suggested that inferior convexity neurons might represent spatial information as well, and conversely dorsolateral prefrontal neurons might represent stimulus identity, particularly after animals have been trained to perform a memory task that requires them to remember both the location and identity of a stimulus (Rao and others 1997; Rainer and others 1998a). These experiments led to the suggestion that information for spatial location and identity might be combined in the prefrontal cortex. Newer studies shed light on this apparent discrepancy. It has now been demonstrated that the object selectivity observed in dorsolateral prefrontal neurons could be accounted for by broad but significant tuning to stimulus shape already present at the level of the posterior parietal cortex (Serenio and Maunsell 1998). Similarly, neurons in the inferior temporal cortex can be highly selective for spatial position (DiCarlo and Maunsell 2003). The selectivity for object and spatial information is therefore a matter of degree (a gradient) in the prefrontal cortex, as a result of the cross-talk between the dorsal and ventral visual pathways upstream and possibly through reciprocal connections between the dorsal and ventral subregions of the prefrontal cortex. In addition, experience-dependent plasticity might endow prefrontal cortical neurons with the ability to encode combined object and spatial information, which could be used to subserve flexible sensorimotor association in behavioral control (Asaad and others 1998; White and Wise 1999).

Within a prefrontal subregion (say, area 46), previous experiments have failed to establish a clear pattern of topographic organization of prefrontal (or posterior pari-

etal) memory fields, unlike the primary visual cortex, in which the entire visual space is represented topographically across the cortical surface. Despite the apparent lack of an overall retinotopic map across the surface of the prefrontal cortex, there is strong evidence for a systematic organization of spatial information at a local level. Localized prefrontal lesions produce behavioral deficits in the execution of mnemonic tasks involving only a restricted area of visual space, typically in the contralateral hemisphere, an effect known as a “mnemonic scotoma” (Funahashi, Bruce, and others 1993). Similarly, chemical inactivation of prefrontal sites produces the inability of monkeys to correctly recall targets appearing at certain spatial locations (Sawaguchi and Goldman-Rakic 1991, 1994), which again is an argument for an organized representation of visual space in the cortex. More recently, simultaneous recordings from closely spaced electrodes have confirmed that neurons in the proximity of each other (laterally separated by less than 0.3 mm) most often represent adjacent spatial locations (Constantinidis and others 2001a). A possible organizational scheme that could account for these results would be for the entire visual hemifield to be represented in repeating, topographically organized cortical modules, perhaps corresponding to the anatomical stripe-like zones of axonal terminations (Levitt and others 1993; Kritzer and Goldman-Rakic 1995; Pucak and others 1996).

Cellular Mechanisms of Spatial Working Memory

The maintenance of neuronal discharge representing remembered information can last for several seconds, much longer than the time constants of single-neuron biophysical processes, and it is therefore thought to be foremost a network function. Single neurons may also be bistable due to intrinsic membrane properties that contribute to network behavior (Camperi and Wang 1998; Wang 2001). Indeed, a recent study of *in vitro* slice recordings reports that individual neurons in the entorhinal cortex are capable of producing graded persistent discharges (Egorov and others 2002). How a network can give rise to sustained activation in the absence of a direct sensory stimulation can be illustrated as follows: A visual stimulus produces activation of primary visual cortical areas, which is ultimately propagated to the prefrontal cortex, where it excites a population of pyramidal neurons tuned to its location in space. These neurons are linked through reciprocal, excitatory synaptic connections so that even when the original stimulus is no longer present, discharges continue to reverberate in the network (Wang 2001). The anatomical organization of the prefrontal cortex in many ways resembles a recurrent neural network (Fig. 4). Axonal projections of pyramidal neurons originating from within the prefrontal cortex (intrinsic projections) as well as from other cortical areas (associational projections) terminate in a precise stripe-like fashion, creating a regular pattern of interdigitated columns, approximately 0.5 mm wide and 2 to 8 mm

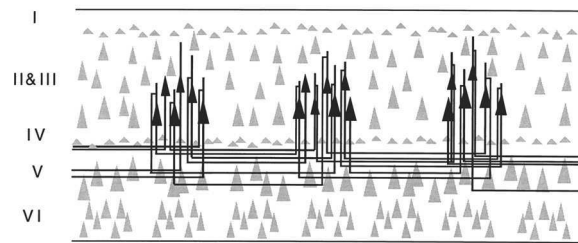


Fig. 4. Schematic diagram illustrating the pattern of connections between prefrontal neurons in the superficial layers. The figure summarizes results of anatomical tracer injection experiments and retrograde labeling. From Kritzer and Goldman-Rakic (1995), with permission.

long (Levitt and others 1993; Kritzer and Goldman-Rakic 1995; Pucak and others 1996). Reciprocal connections between neurons in such stripes create the anatomical substrate of a recurrent network.

Biophysically realistic computational modeling has shown that such recurrent networks can give rise to location-specific, persistent discharges, as illustrated in Figure 5 (Camperi and Wang 1998; Compte and others 2000; Gutkin and others 2001; Tegner and others 2002; Renart and others 2003a, Wang, Tegner, and others 2004). The central conceptual idea that emerges from these studies is that a spatial working memory network should be bistable between a spontaneous resting state and a continuous family of spatially localized persistent firing patterns (each encoding and storing a spatial location as an analog quantity). The appearance of the transient cue stimulus switches the network from the resting state to a memory state representing a spatial location, and the feedback signal or reward that indicates the end of the trial erases the memory trace and switches the network back to its resting state. It was realized that the instantiation of a continuum of persistent states requires the cortical network to be functionally homogeneous, that is, to be characterized by identical cellular and synaptic properties between its neurons. Such homogeneity is unrealistic for real-neuron networks but could be effectively achieved by biological mechanisms such as homeostatic regulations that could scale the synaptic inputs of neurons with different rates of activity (Renart and others 2003a). It is also worth noting that the continuous nature of spatial working memory has not been scrutinized in experimental studies, in which only a limited number (typically eight) of spatial cues are used. It would be desirable in future psychophysical and electrophysiological studies to rigorously test the analog character of spatial working memory by using an arbitrary set of cues.

Understanding the dynamical behavior of strongly recurrent cortical microcircuits, including those presumably required for persistent activity, presents experimental and theoretical challenges. In biophysically realistic models, such a neural network is prone to dynamical destabilization leading to either uncontrolled spike discharges at very high rates, if excitatory reverberation is

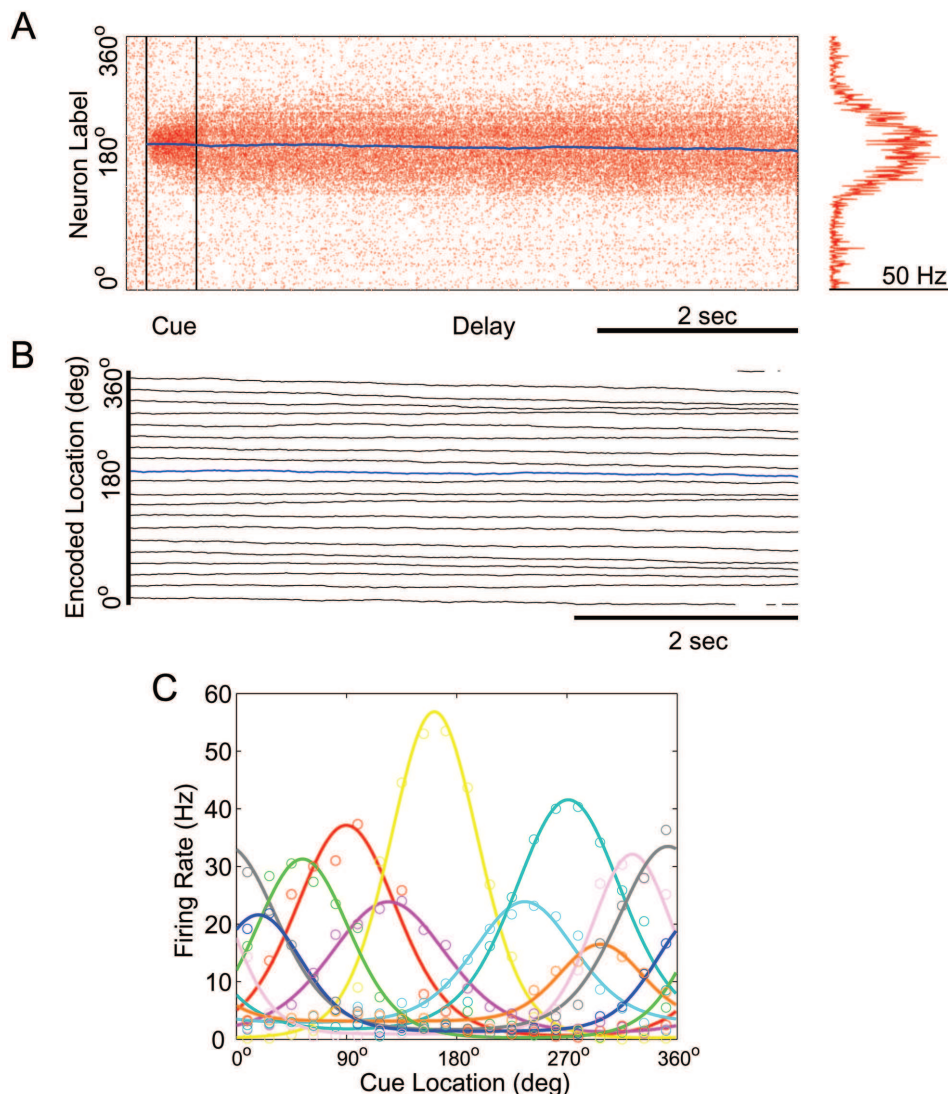


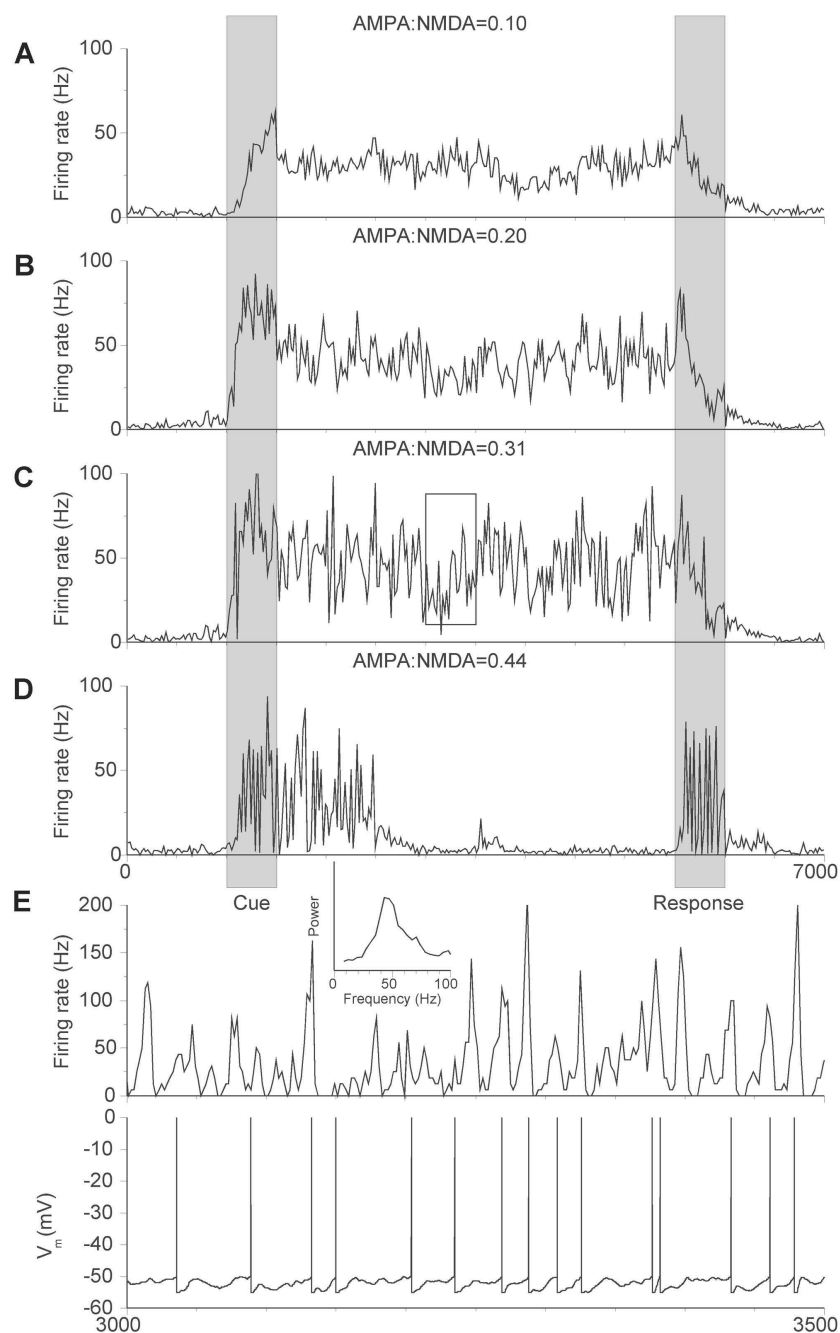
Fig. 5. A, Model simulation of the delayed oculomotor response experiment (Figs. 1, 2). Spatiotemporal activity pattern of the pyramidal cell population. A dot at position (t, θ) represents a spike fired at time t by a cell with preferred cue location at direction θ . The blue line represents the time evolution of the peak location of the persistent activity pattern. Right, localized activity profile during the mnemonic delay period. B, Temporal evolution of the peak location of memory activity in 20 trials with transient stimuli at different locations. The memory of the initial cue is well preserved during the 6-second delay period. C, Tuning curves from 10 neurons. The firing rate of each cell for the 20 stimuli in (B) is shown as open circles. Adapted with permission from Renart, Song, and others (2003).

not counterbalanced by rate-control mechanisms, or to excessive synchronous oscillations that may be disruptive to the maintenance of persistent activity. Working memory function was found to be particularly stable (Fig. 6) when excitatory reverberation was characterized by a fairly slow time course, for example, when synaptic transmission at intrinsic connections incorporated a large NMDA component (Wang 1999; Compte and others 2000; Wang 2001; Tegner 2002; Renart and others 2003b). NMDA receptor-mediated recurrent excitation may also be especially important for the robustness of short-term memory that encodes continuous quantities, such as spatial location or eye position (Seung and others 2000). The voltage dependence of NMDA channels may contribute to stimulus selectivity of a persistent firing pattern (Lisman and others 1998).

Temporal Dynamics of Persistent Activity

An interesting question is whether working memory is stored mainly in the elevated firing rate of a subpopulation of neurons or depends on more complicated temporal discharge patterns and cross-neuronal synchrony. In principle, stimulus information can be conveyed by different temporal patterns of activity without any variations in the average firing rate of a neuron. In fact, under some conditions, the averaged responses of the entire prefrontal population may vary little between a spontaneous and an active memory state (Shafi and others 2003). Such a result must be treated with caution because different classes of excitatory and inhibitory neurons play distinct roles during working memory and may increase or decrease their firing rates in response to a stimulus (further discussed below). Pooling all respons-

Fig. 6. Stability of persistent activity as a function of the AMPA:NMDA ratio at the recurrent excitatory synapses. *A–D*, Temporal course of the average firing rate across a subpopulation of cells selective to the presented transient input, for different levels of the AMPA:NMDA ratio. As the ratio is increased, oscillations of a progressively larger amplitude develop during the delay period, which eventually destabilize the persistent activity state. *E*, Snapshot of the activity of the network in (*C*) between 3 and 3.5 seconds. *Top*, Average network activity. *Bottom*, Intracellular voltage trace of a single neuron. *Inset*, Power spectrum of the average activity of the network, showing a peak in the γ (40 Hz) frequency range. Persistent activity is stable even in the presence of synchronous oscillations. Adapted with permission from Renart, Brunel, and others (2003).



es together may therefore be misleading and does not argue against a rate code.

The question of temporal coding in the spiking patterns of prefrontal cortical neurons was recently addressed by a temporal-statistics analysis (Compte and others 2003). The results revealed that most single units of monkey prefrontal cortex during the mnemonic delay can be well approximated by a purely random, Poisson process that contains no systematic temporal patterns of firing. A smaller proportion of prefrontal neurons exhibit burst discharges. The coefficient of variation of prefrontal interspike intervals is typically near or larger than

1, indicating a high degree of irregularity of delay period neural discharges, consistent with a network mechanism for persistent activity sustained by stochastic synaptic bombardments. Rhythmicity is rarely visible in spike trains of single prefrontal cells during working memory (Compte and others 2003). It is possible that the network might display coherent oscillations detectable only in the population level, for example, reflected in local field potential measurements summing responses from large numbers of neurons, whereas single cells are highly stochastic (Pesaran and others 2002; Brunel and Wang 2003). Regardless of whether oscillations are

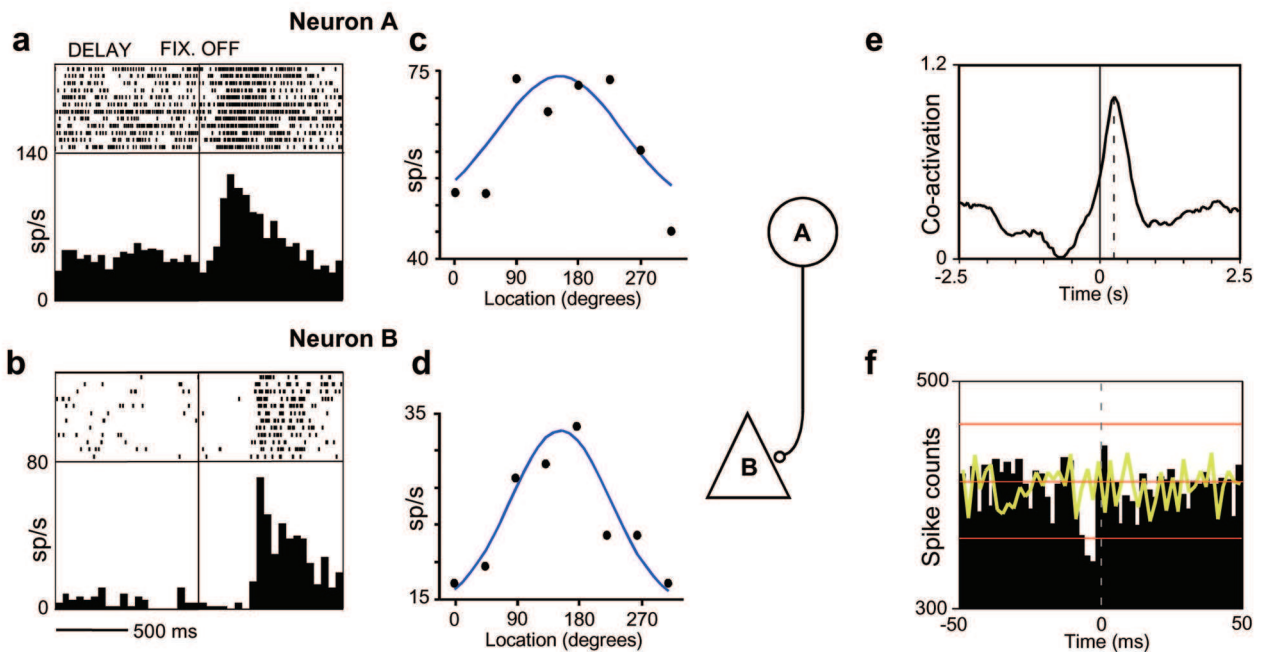


Fig. 7. An example of an inhibitory interaction detected through simultaneous recordings from two microelectrodes. *a–b*, Responses of two neurons recorded 0.3 mm apart. *c–d*, The neurons' spatial tuning, computed in the saccade period. *e*, Coactivation function, indicating the time lag that produced the maximum temporal overlap of the two neurons' responses, shows a 250-ms lag. *f*, Cross-correlation histogram revealing a trough, offset from time zero, consistent with direct inhibition from neuron A onto neuron B. From Constantinidis and others (2002), with permission.

present, there is little doubt that synchronous firing arises between prefrontal neurons during working memory, as shown by simultaneously recorded pairs of single units from behaving monkeys (Constantinidis and Goldman-Rakic 2002). In a rodent study of delayed spatial task, simultaneous recording from many (>20) single prefrontal neurons by multiple-contact electrodes (tetrodes) suggests that working memory could be maintained in a spatio-temporal firing pattern propagating around the network, in which any given cell needs only to fire at a high rate episodically rather than tonically throughout the memory period (Baeg and others 2003). This result, if confirmed by future work especially with primates, would offer a new picture about the dynamic organization of a working memory network. Progress in this direction necessitates the development of techniques for simultaneous recording of large numbers of single neurons or imaging of spatio-temporal firing patterns in the working memory cortical network of behaving monkeys.

Memory Fields Are Shaped by Inhibition

Prefrontal cortical neurons have large, often bilateral memory fields, which are, however, spatially distinct; they are centered at specific spatial locations and are typically flanked by inhibitory surrounds. Inhibition is difficult to detect in neurophysiological recordings, as the baseline discharge rate of pyramidal neurons is low and the inhibitory effect of a remembered stimulus can

easily go unnoticed. Nonetheless, clear examples of inhibitory fields away from a neuron's preferred location have been demonstrated in neurophysiological recordings (Funahashi and others 1989). A recent population analysis of a large neuronal database confirmed that the discharge rate after the presentation of a memorandum at a location diametric to a neuron's preferred location has an overall suppressive effect below the level of background activity (Constantinidis and Goldman-Rakic 2002). A direct demonstration of the role of inhibition has been provided by micro-iontophoretic application of GABA antagonists in vivo. Such negation of inhibitory inputs expands the size of the neurons' memory fields, unmasking excitatory zones that were suppressed during normal neuronal function (Rao and others 2000). Inhibitory interactions have also been demonstrated by means of simultaneous recordings and cross-correlation analysis, which also indicated that interneurons inhibit pyramidal neurons active at different time points of the behavioral task (Fig. 7), or exhibiting dissimilar spatial receptive fields (Constantinidis and others 2002).

Pyramidal and nonpyramidal neurons tend to exhibit distinct action potential waveforms and rates of discharges, making it possible to differentiate them in extracellular recordings in vivo (Wilson and others 1994; Constantinidis and Goldman-Rakic 2002). It has been thus possible to examine the properties of inhibitory neurons themselves during the maintenance of information in memory. Interneurons receive inputs from nearby pyramidal neurons, and they too exhibit sus-

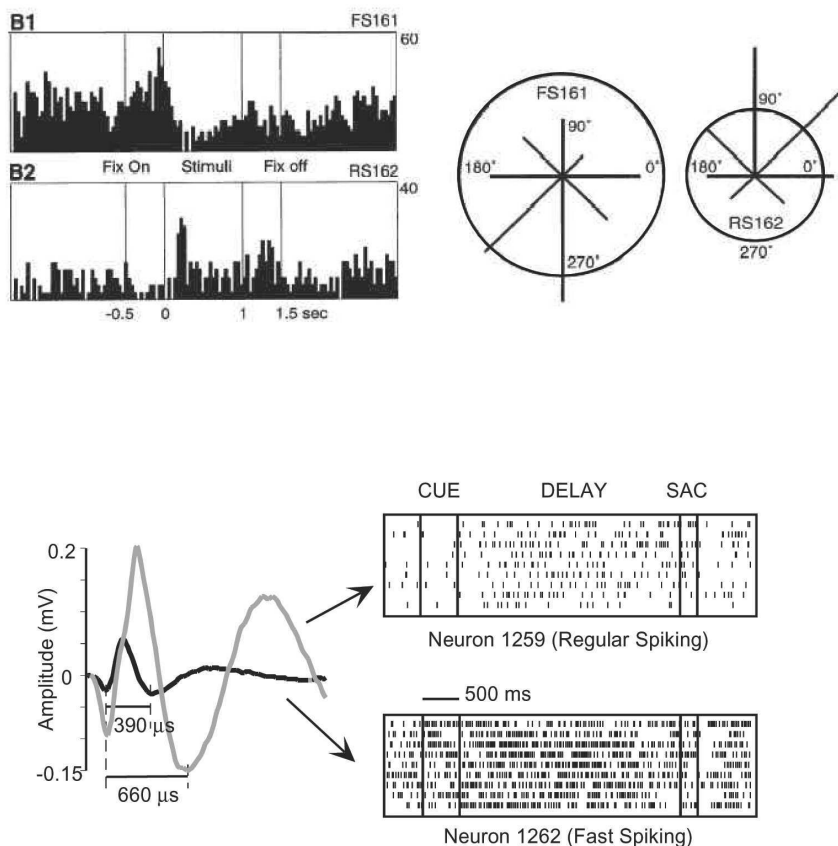


Fig. 8. Responses of putative inhibitory interneurons during the execution of the delayed response task. *Top*, inverted tuning of an interneuron (top histogram and left polar plot) relative to a pyramidal neuron (bottom histogram and right polar plot), recorded in sequence, 0.2 mm apart. The activity of the interneuron is depressed below the baseline during the delay period following presentation of the cue that evokes the best response of the pyramidal neuron (from Wilson and others 1994, with permission). *Bottom*, elevated firing rate is observed for an interneuron (bottom raster plot) and a pyramidal neuron (top raster plot) after presentation of the cue in a location that evoked the best response for both neurons, which were recorded simultaneously from electrodes 0.2 mm apart. Data from Constantinidis and Goldman-Rakic (2002).

tained responses and spatially tuned memory fields. Initial reports from sequential microelectrode recordings within 0.4 mm of each other suggested that interneurons display inverted tuning relative to pyramidal neurons (Fig. 8); interneuron discharges were found to decrease below the baseline during the maintenance in memory of a stimulus in the preferred location of nearby pyramidal neurons (Wilson and others 1994). Later results, based on simultaneous recordings from the same electrode or separate electrodes 0.2 to 0.3 mm apart, indicated that nearby pyramidal neurons and interneurons are tuned to similar spatial locations and demonstrate persistent elevation of neuronal activity (Rao and others 1999; Constantinidis and Goldman-Rakic 2002). Pyramidal cells and interneurons recorded at longer distances from each other may be tuned to opposing spatial locations. Neurons with suppressed responses below the baseline now appear to constitute a distinct subclass of interneurons, as suggested by computational modeling studies. Spatially tuned, sustained activity can be achieved by a network architecture in which nearby pyramidal cells and interneurons show similar selectivity, whereas further separated pyramidal cells and interneurons display opposite tuning (Fig. 9A). A diversity of inhibitory neurons within each cortical module cooperate to provide synaptic inhibition, with different inhibitory neurons

exhibiting different patterns of excitation. A most recent study proposes a theoretical framework for how three major subclasses of interneurons conspire to subserve spatial working memory (Fig. 9C). In this model, inverted tuning, which was originally thought as a defining property of prefrontal interneurons, is displayed by a distinct subclass of interneurons, possibly corresponding to calbindin immunoreactive cells (Wang, Tegner, and others 2004).

The importance of inhibition in the spatial working memory can be illustrated by noting that localized persistent activity can be achieved without any recurrent excitation at all. Instead, significant background external excitation that drives all neurons to fire at relatively high rates, offset by feedback cross-directional inhibition, can sculpture the tuning of activity in a network, allowing interneurons to suppress those excitatory cells that are selective for the cue stimulus (Fig. 9B). Such a network architecture may be the basis of persistent activity that encodes the head direction of an animal during spatial navigation (Sharp and others 2001).

Spatial Working Memory and Mental Illness

Cognitive processes that involve spatial working memory are compromised by a number of mental illnesses,

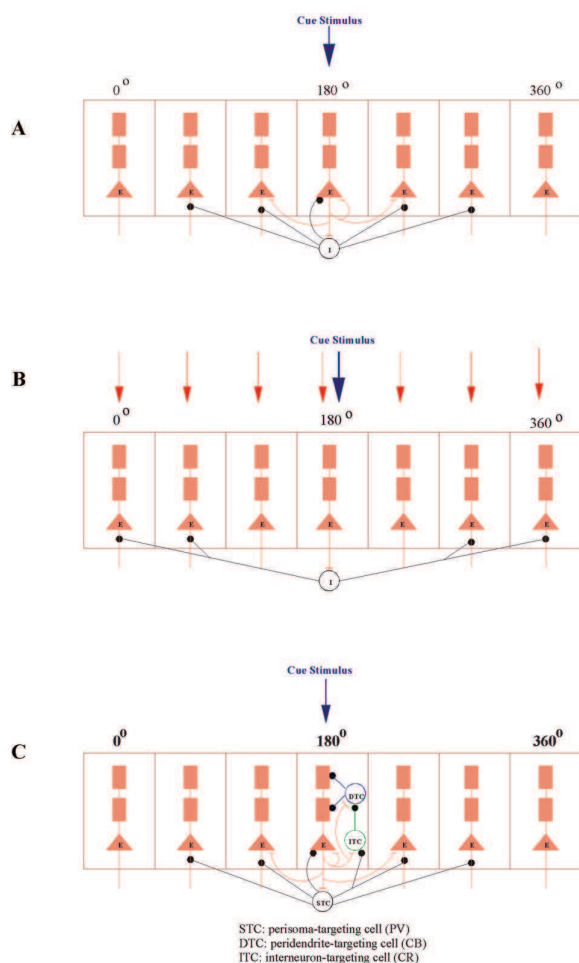


Fig. 9. Three network circuit schemes for spatially tuned persistent activity patterns. *A*, Local recurrent excitation sustains elevated firing in a group of cells, whereas iso-directional inhibition of broader projection extent suppresses excitatory cells on the flanks of the hill of activity. *B*, Tonic external inputs provide uniform drive to the entire excitatory cell population, and cross-directional feedback inhibition gives rise to spatial localization of persistent activity. *C*, Coordinated operation by three subtypes of inhibitory cells in a spatial working memory network. Perisoma-targeting (parvalbumin-containing [PV]) cells provide lateral inhibition, as in (*A*). Within a column, calbindin-containing (CB) interneurons target the dendrites of excitatory neurons, whereas calretinin-containing (CR) interneurons preferentially project to CB cells. Excitation of a group of excitatory cells recruits locally CR neurons, which sends enhanced inhibition to CB neurons, leading to dendritic disinhibition of the same excitatory cells. (*C*) is taken from Wang, Tegner, and others (2004), with permission.

including schizophrenia (Park and Holzman 1992), obsessive compulsive disorder (Purcell and others 1998), major depressive disorder (Murphy and others 2003), and chronic alcoholism (Sullivan and others 1993).

Spatial working memory impairment has been most extensively studied in schizophrenia. Positive symptoms of schizophrenia include delusions, hallucinations, thought disorders, and attentional impairments, and they respond well to typical antipsychotic drugs. Negative

symptoms (absence of normal traits) include affective and motivational deficits, emotional and social withdrawal, disorganized speech, and anhedonia. Schizophrenic patients exhibit impaired performance in spatial working memory tasks that involve eye or manual movements toward the remembered direction of a visual target presented a few seconds earlier or that require them to keep track of the locations of visual stimuli presented or sampled in sequence (Park and others 1995; Pantelis and others 1997). Schizophrenic patients are similarly impaired in antisaccade tasks, requiring an eye movement in the direction opposite to a visual target, and to smooth-pursuit eye movements, tracking a moving visual stimulus (Ettinger and others 2004; Reuter and others 2004). Impairment of spatial working memory performance has been observed in patients with both negative and positive symptoms of schizophrenia, including those with psychosis, those who are medicated and unmedicated, those in the acute phase of illness or in relapse, and even in undiagnosed relatives of schizophrenic patients (Park and others 1995; Carter and others 1996; Park and others 1999; Wood and others 2003). Given the strong linkage between schizophrenia and spatial working memory, performance in spatial working memory tasks has been considered a potential predictive indicator and a marker for the genetic liability to the disease (Fuller and others 2002; Gasperoni and others 2003; McGorry 2003; Niendam and others 2003). The link between working memory and schizophrenia may be the consequence of prefrontal malfunction giving origin to both diminished ability to maintain information in memory and a host of other cognitive distortions. On the other hand, working memory impairment may be a contributing cause of the cognitive defects (Goldman-Rakic 1994).

Why and how spatial working memory is impaired in schizophrenic patients cannot be truly answered at a fundamental level without a deep understanding of the underlying cellular and circuit mechanisms. For example, when the inhibitory circuit organization in the prefrontal cortex is elucidated, it will become clear why particular kinds of defects in cortical interneurons cause behavioral impairments (Lewis and others 2003). Another important example comes from recent progress in studies of dopamine modulation of spatial working memory. Schizophrenia is associated with changes in dopamine innervation and action. The fact that the frontal lobe receives a much more prominent dopaminergic innervation compared to the parietal or occipital areas involved with spatial stimulus representation has also led to speculation that unique aspects of memory maintenance associated with the prefrontal cortex are dopamine dependent.

The effects of dopamine are complex and in many ways remain unclear. Direct injections of dopamine agonists in the prefrontal cortex of monkeys performing a memory task were shown to degrade behavioral performance at the spatial location encoded by the injected site, suggesting a possible inhibitory role of dopamine (Sawaguchi and Goldman-Rakic 1991). However, finer

studies of dopamine agonist micro-iontophoresis in the prefrontal cortex revealed excitatory effects at low concentrations and a dose-dependent effect (Williams and Goldman-Rakic 1995). Dopamine application in prefrontal slice preparations increases the excitability of pyramidal neurons, at least in the superficial cortical layers (Henze and others 2000). On the other hand, dopamine can decrease the presynaptic efficacy of neurotransmission onto pyramidal neurons in layer 5 (Gao and others 2001), and it may even have opposite effects on different synapses of the same neuron: It has been shown to increase the efficacy of inhibitory synapses targeting the dendritic tree, while decreasing the inhibitory effect of perisomatic synapses (Gao and others 2003). Functionally, using micro-iontophoresis in behaving monkeys, it was found that different subtypes of dopamine receptors mediate distinct effects on neural activities responsible for different epochs of the task: D1 receptors affect memory storage during the delay, whereas D2 receptors modulate the neural activity during the behavioral response at the end of the trial (Wang, Vijayraghavan, and others 2004). As a result of the multiple actions of dopamine, episodic hypo- and hyperactivity of the prefrontal cortex may coexist in the same schizophrenic patient in the course of the illness (Seamans and Yang 2004).

Although the effects of dopamine are complex, they are consistent with the idea that dopamine increases the signal-to-noise ratio of activity representing a remembered stimulus (Camperi and Wang 1998; Durstewitz and others 2000; Brunel and Wang 2001; Cohen and others 2002). Enhancement of dendritic inhibition during the maintenance of memory could be an effective mechanism of filtering distracting, sensory information (Wang and others 2004).

Conclusions

In summary, thanks to a highly interdisciplinary approach combining *in vitro* and *in vivo* electrophysiology and computational modeling, there has been significant progress in our understanding of the cortical basis of spatial working memory. However, we are still faced with many challenges ahead. Functional questions about the organization of the prefrontal cortex and its spatial working memory function remain unanswered. Little is known about the precise role of NMDA receptors to excitatory reverberation, the respective roles of and the interplay between prefrontal and parietal cortices, and the continuous nature of spatial encoding at the neural level. Progress on these issues will help to arrive at a deep understanding of the circuit mechanisms of spatial working memory and shed light into the cellular origins of memory impairments associated with schizophrenia and other mental disorders.

References

- Andersen RA, Essick GK, Siegel RM. 1987. Neurons of area 7 activated by both visual stimuli and oculomotor behavior. *Exp Brain Res* 67:316–22.
- Asaad WF, Rainer G, Miller EK. 1998. Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21:1399–407.
- Asaad WF, Rainer G, Miller EK. 2000. Task-specific neural activity in the primate prefrontal cortex. *J Neurophysiol* 84:451–9.
- Azuma M, Suzuki H. 1984. Properties and distribution of auditory neurons in the dorsolateral prefrontal cortex of the alert monkey. *Brain Res* 298:343–6.
- Baddeley A. 1992. Working memory. *Science* 255:556–9.
- Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW. 2003. Dynamics of population code for working memory in the prefrontal cortex. *Neuron* 40:177–88.
- Bisley JW, Zaksas D, Droll J, Pasternak T. 2004. Activity of neurons in cortical area MT during a memory for motion task. *J Neurophysiol* 91:286–300.
- Bodner M, Kroger J, Fuster JM. 1996. Auditory memory cells in dorsolateral prefrontal cortex. *Neuroreport* 7:1905–8.
- Brunel N, Wang XJ. 2001. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci* 11:63–85.
- Brunel N, Wang XJ. 2003. What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *J Neurophysiol* 90:415–30.
- Bunge SA, Ochsner KN, Desmond JE, Glover GH, Gabrieli JD. 2001. Prefrontal regions involved in keeping information in and out of mind. *Brain* 124:2074–86.
- Camperi M, Wang XJ. 1998. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J Comput Neurosci* 5:383–405.
- Carter CS, Robertson LC, Nordahl TE, Kraft L, Chaderjian M, Oshora-Celaya L. 1996. Spatial working memory deficits and their relationship to negative symptoms in unmedicated schizophrenic patients. *Biol Psychiatry* 40:930–2.
- Chafee MV, Goldman-Rakic PS. 1998. Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *J Neurophysiol* 79:2919–40.
- Cohen JD, Braver TS, Brown JW. 2002. Computational perspectives on dopamine function in prefrontal cortex. *Curr Opin in Neurobiol* 12:223–229.
- Compte A, Brunel N, Goldman-Rakic PS, Wang XJ. 2000. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10:910–23.
- Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, and others. 2003. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol* 28:28.
- Constantinidis C, Franowicz MN, Goldman-Rakic PS. 2001a. Coding specificity in cortical microcircuits: a multiple electrode analysis of primate prefrontal cortex. *J Neurosci* 21:3646–55.
- Constantinidis C, Franowicz MN, Goldman-Rakic PS. 2001b. The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nat Neurosci* 4:311–6.
- Constantinidis C, Goldman-Rakic PS. 2002. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *J Neurophysiol* 88:3487–97.
- Constantinidis C, Steinmetz MA. 1996. Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *J Neurophysiol* 76:1352–5.
- Constantinidis C, Steinmetz MA. 2001. Neuronal responses in area 7a to multiple stimulus displays: II. Responses are suppressed at the cued location. *Cereb Cortex* 11:592–7.
- Constantinidis C, Williams GV, Goldman-Rakic PS. 2002. A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nat Neurosci* 5:175–80.
- Cornette L, Dupont P, Orban GA. 2002. The neural substrate of orientation short-term memory and resistance to distractor items. *Eur J Neurosci* 15:165–75.
- Courtney SM, Ungerleider LG, Keil K, Haxby JV. 1997. Transient and sustained activity in a distributed neural system for human working memory. *Nature* 386:608–11.
- Curtis CE, D'Esposito M. 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn Sci* 7:415–23.

- di Pellegrino G, Wise SP. 1993a. Effects of attention on visuomotor activity in the premotor and prefrontal cortex of a primate. *Somatosens Mot Res* 10:245–62.
- di Pellegrino G, Wise SP. 1993b. Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *J Neurosci* 13:1227–43.
- DiCarlo JJ, Maunsell JH. 2003. Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *J Neurophysiol* 89:3264–78.
- Durstewitz D, Seamans JK, Sejnowski TJ. 2000. Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. *J Neurophysiol* 83:1733–50.
- Egorov AV, Hamam BN, Franssen E, Hasselmo ME, Alonso AA. 2002. Graded persistent activity in entorhinal cortex neurons. *Nature* 420:173–8.
- Ettinger U, Kumari V, Crawford TJ, Corr PJ, Das M, Zachariah E, and others. 2004. Smooth pursuit and antisaccade eye movements in siblings discordant for schizophrenia. *J Psychiatr Res* 38:177–84.
- Everling S, Tinsley CJ, Gaffan D, Duncan J. 2002. Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nat Neurosci* 5:671–6.
- Fuller R, Nopoulos P, Arndt S, O'Leary D, Ho BC, Andreasen NC. 2002. Longitudinal assessment of premorbid cognitive functioning in patients with schizophrenia through examination of standardized scholastic test performance. *Am J Psychiatry* 159:1183–9.
- Funahashi S, Bruce CJ, Goldman-Rakic PS. 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J Neurophysiol* 61:331–49.
- Funahashi S, Bruce CJ, Goldman-Rakic PS. 1993. Dorsolateral prefrontal lesions and oculomotor delayed-response performance: evidence for mnemonic "scotomas." *J Neurosci* 13:1479–97.
- Funahashi S, Chafee MV, Goldman-Rakic PS. 1993. Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature* 365:753–6.
- Fuster JM, Alexander GE. 1971. Neuron activity related to short-term memory. *Science* 173:652–4.
- Fuster JM, Alexander GE. 1973. Firing changes in cells of the nucleus medialis dorsalis associated with delayed response behavior. *Brain Res* 61:79–91.
- Gao WJ, Krimer LS, Goldman-Rakic PS. 2001. Presynaptic regulation of recurrent excitation by D1 receptors in prefrontal circuits. *Proc Natl Acad Sci U S A* 98:295–300.
- Gao WJ, Wang Y, Goldman-Rakic PS. 2003. Dopamine modulation of perisomatic and peridendritic inhibition in prefrontal cortex. *J Neurosci* 23:1622–30.
- Gasperoni TL, Ekelund J, Huttunen M, Palmer CG, Tuulio-Henriksson A, Lonnqvist J, and others. 2003. Genetic linkage and association between chromosome 1q and working memory function in schizophrenia. *Am J Med Genet* 116B:8–16.
- Glimcher PW, Sparks DL. 1992. Movement selection in advance of action in the superior colliculus. *Nature* 355:542–5.
- Gnadt JW, Andersen RA. 1988. Memory related motor planning activity in posterior parietal cortex of macaque. *Exp Brain Res* 70:216–20.
- Goldman-Rakic PS. 1987. Circuitry of the prefrontal cortex and the regulation of behavior by representational knowledge. In: Plum F, Mountcastle VB, editors. *Handbook of physiology*. Bethesda (MD): American Physiological Society. p 373–417.
- Goldman-Rakic PS. 1994. Working memory dysfunction in schizophrenia. *J Neuropsychiatry Clin Neurosci* 6:348–57.
- Grunewald A, Linden JF, Andersen RA. 1999. Responses to auditory stimuli in macaque lateral intraparietal area: I. Effects of training. *J Neurophysiol* 82:330–42.
- Gutkin BS, Laing CR, Colby CL, Chow CC, Ermentrout GB. 2001. Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. *J Comput Neurosci* 11:121–34.
- Henze DA, Gonzalez-Burgos GR, Urban NN, Lewis DA, Barrionuevo G. 2000. Dopamine increases excitability of pyramidal neurons in primate prefrontal cortex. *J Neurophysiol* 84:2799–809.
- Hikosaka O, Sakamoto M, Usui S. 1989. Functional properties of monkey caudate neurons: I. Activities related to saccadic eye movements. *J Neurophysiol* 61:780–98.
- Hikosaka O, Wurtz RH. 1983. Visual and oculomotor functions of monkey substantia nigra pars reticulata: III. Memory-contingent visual and saccade responses. *J Neurophysiol* 49:1268–84.
- Jacobsen CF. 1936. Studies of cerebral function in primates. *Comp Psychol Monogr* 13:1–68.
- Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA. 1993. Spatial working memory in humans as revealed by PET. *Nature* 363:623–5.
- Kimura M, Aosaki T, Hu Y, Ishida A, Watanabe K. 1992. Activity of primate putamen neurons is selective to the mode of voluntary movement: visually guided, self-initiated or memory-guided. *Exp Brain Res* 89:473–7.
- Kritzer MF, Goldman-Rakic PS. 1995. Intrinsic circuit organization of the major layers and sublayers of the dorsolateral prefrontal cortex in the rhesus monkey. *J Comp Neurol* 359:131–43.
- Kubota K, Niki H. 1971. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J Neurophysiol* 34:337–47.
- Leon MI, Shadlen MN. 1999. Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–25.
- Levitt JB, Lewis DA, Yoshioka T, Lund JS. 1993. Topography of pyramidal neuron intrinsic connections in macaque monkey prefrontal cortex (areas 9 and 46). *J Comp Neurol* 338:360–76.
- Lewis DA, Volk DW, Hashimoto T. 2003. Selective alterations in prefrontal cortical GABA neurotransmission in schizophrenia: a novel target for the treatment of working memory dysfunction. *Psychopharmacology (Berl)* 9:9.
- Lisman JE, Fellous JM, Wang XJ. 1998. A role for NMDA-receptor channels in working memory. *Nat Neurosci* 1:273–5.
- Marshuetz C, Smith EE, Jonides J, DeGutis J, Chenevert TL. 2000. Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J Cogn Neurosci* 12:130–44.
- Mazzoni P, Bracewell RM, Barash S, Andersen RA. 1996. Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *J Neurophysiol* 75:1233–41.
- McGorry PD. 2003. Spatial working memory ability is a marker of risk-for-psychosis. *Psychol Med* 33:1239–47.
- Milner B. 1963. Effects of different brain lesions on card sorting. *Arch Neurol* 9:100–10.
- Munk MH, Linden DE, Muckli L, Lanfermann H, Zanella FE, Singer W, and others. 2002. Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. *Cereb Cortex* 12:866–76.
- Murphy FC, Michael A, Robbins TW, Sahakian BJ. 2003. Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. *Psychol Med* 33:455–67.
- Nakamura K, Colby CL. 2000. Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *J Neurophysiol* 84:677–92.
- Niendam TA, Bearden CE, Rosso IM, Sanchez LE, Hadley T, Nuechterlein KH, and others. 2003. A prospective study of childhood neurocognitive functioning in schizophrenic patients and their siblings. *Am J Psychiatry* 160:2060–2.
- Niki H, Watanabe M. 1976. Prefrontal unit activity and delayed response: relation to cue location versus direction of response. *Brain Res* 105:79–88.
- Owen AM, Stern CE, Look RB, Tracey I, Rosen BR, Petrides M. 1998. Functional organization of spatial and nonspatial working memory processing within the human lateral frontal cortex. *Proc Natl Acad Sci U S A* 95:7721–6.
- Pantelis C, Barnes TR, Nelson HE, Tanner S, Weatherley L, Owen AM, and others. 1997. Frontal-striatal cognitive deficits in patients with chronic schizophrenia. *Brain* 120:1823–43.
- Park S, Holzman PS. 1992. Schizophrenics show spatial working memory deficits. *Arch Gen Psychiatry* 49:975–82.
- Park S, Holzman PS, Goldman-Rakic PS. 1995. Spatial working memory deficits in the relatives of schizophrenic patients. *Arch Gen Psychiatry* 52:821–8.
- Park S, Puschel J, Sauter BH, Rentsch M, Hell D. 1999. Spatial working memory deficits and clinical symptoms in schizophrenia: a 4-month follow-up study. *Biol Psychiatry* 46:392–400.

- Pesaran B, Pezaris JS, Sahani M, Mitra PP, Andersen RA. 2002. Temporal structure in neuronal activity during working memory in macaque parietal cortex. *Nat Neurosci* 5:805–11.
- Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400:233–8.
- Powell KD, Goldberg ME. 2000. Response of neurons in the lateral intraparietal area to a distractor flashed during the delay period of a memory-guided saccade. *J Neurophysiol* 84:301–10.
- Pucak ML, Levitt JB, Lund JS, Lewis DA. 1996. Patterns of intrinsic and associational circuitry in monkey prefrontal cortex. *J Comp Neurol* 376:614–30.
- Purcell R, Maruff P, Kyrios M, Pantelis C. 1998. Neuropsychological deficits in obsessive-compulsive disorder: a comparison with unipolar depression, panic disorder, and normal controls. *Arch Gen Psychiatry* 55:415–23.
- Quintana J, Fuster JM. 1992. Mnemonic and predictive functions of cortical neurons in a memory task. *Neuroreport* 3:721–4.
- Rainer G, Asaad WF, Miller EK. 1998a. Memory fields of neurons in the primate prefrontal cortex. *Proc Natl Acad Sci U S A* 95:15008–13.
- Rainer G, Asaad WF, Miller EK. 1998b. Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature* 393:577–9.
- Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. *Science* 276:821–4.
- Rao SG, Williams GV, Goldman-Rakic PS. 1999. Isodirectional tuning of adjacent interneurons and pyramidal cells during working memory: evidence for microcolumnar organization in PFC. *J Neurophysiol* 81:1903–16.
- Rao SG, Williams GV, Goldman-Rakic PS. 2000. Destruction and creation of spatial tuning by disinhibition: GABA(A) blockade of prefrontal cortical neurons engaged by working memory. *J Neurosci* 20:485–94.
- Renart A, Brunel N, Wang X-J. 2003. Mean-field theory of recurrent cortical networks: working memory circuits with irregularly spiking neurons. In: Feng J, editor. *Computational neuroscience: a comprehensive approach*. Boca Raton (FL): CRC Press. p 432–90.
- Renart A, Song P, Wang X-J. 2003. Robust spatial working memory in a heterogeneous network model with homeostatic synaptic scaling. *Neuron* 38:473–85.
- Reuter B, Kathmann N, Ettinger U, Kumari V, Crawford TJ, Corr PJ, and others. 2004. Using saccade tasks as a tool to analyze executive dysfunctions in schizophrenia. *Acta Psychol (Amst)* 115:255–69.
- Robinson DL, Petersen SE, Keys W. 1986. Saccade-related and visual activities in the pulvinar nuclei of the behaving rhesus monkey. *Exp Brain Res* 62:625–34.
- Rolls ET. 1999. Spatial view cells and the representation of place in the primate hippocampus. *Hippocampus* 9:467–80.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–6.
- Rowe JB, Passingham RE. 2001. Working memory for location and time: activity in prefrontal area 46 relates to selection rather than maintenance in memory. *Neuroimage* 14:77–86.
- Russo GS, Bruce CJ. 1996. Neurons in the supplementary eye field of rhesus monkeys code visual targets and saccadic eye movements in an oculocentric coordinate system. *J Neurophysiol* 76:825–48.
- Sakai K, Rowe JB, Passingham RE. 2002. Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nat Neurosci* 5:479–84.
- Sawaguchi T, Goldman-Rakic PS. 1991. D1 dopamine receptors in prefrontal cortex: involvement in working memory. *Science* 251:947–50.
- Sawaguchi T, Goldman-Rakic PS. 1994. The role of D1-dopamine receptor in working memory: local injections of dopamine antagonists into the prefrontal cortex of rhesus monkeys performing an oculomotor delayed-response task. *J Neurophysiol* 71:515–28.
- Sawaguchi T, Yamane I. 1999. Properties of delay-period neuronal activity in the monkey dorsolateral prefrontal cortex during a spatial delayed matching-to-sample task. *J Neurophysiol* 82:2070–80.
- Seamans JK, Yang CR. 2004. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog Neurobiol* 74:1–58.
- Sereno AB, Maunsell JH. 1998. Shape selectivity in primate lateral intraparietal cortex. *Nature* 395:500–3.
- Seung HS, Lee DD, Reis BY, Tank DW. 2000. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* 26:259–71.
- Shafi M, Bodner M, Zhou Y, Fuster JM. 2003. Measurements of neuronal variability during working memory: implications for theoretical and computational models. *Soc Neurosci Abstr* 33:518.17.
- Sharp PE, Blair HT, Cho J. 2001. The anatomical and computational basis of the rat head-direction cell signal. *Trends Neurosci* 24:289–94.
- Sommer MA, Wurtz RH. 2004. What the brainstem tells the frontal cortex I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *J Neurophysiol* 91:1381–402.
- Stern CE, Sherman SJ, Kirchhoff BA, Hasselmo ME. 2001. Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus* 11:337–46.
- Sullivan EV, Mathalon DH, Zipursky RB, Kerstein-Tucker Z, Knight RT, Pfefferbaum A. 1993. Factors of the Wisconsin Card Sorting Test as measures of frontal-lobe function in schizophrenia and in chronic alcoholism. *Psychiatry Res* 46:175–99.
- Super H, Spekreijse H, Lamme VA. 2001. A neural correlate of working memory in the monkey primary visual cortex. *Science* 293:120–4.
- Suzuki WA, Miller EK, Desimone R. 1997. Object and place memory in the macaque entorhinal cortex. *J Neurophysiol* 78:1062–81.
- Takeda K, Funahashi S. 2002. Prefrontal task-related activity representing visual cue location or saccade direction in spatial working memory tasks. *J Neurophysiol* 87:567–88.
- Tegner J, Compte A, Wang XJ. 2002. The dynamical stability of reverberatory neural circuits. *Biol Cybern* 87:471–81.
- Ungerleider LG, Courtney SM, Haxby JV. 1998. A neural system for human visual working memory. *Proc Natl Acad Sci U S A* 95:883–90.
- Vaadia E, Benson DA, Hienz RD, Goldstein MH Jr. 1986. Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. *J Neurophysiol* 56:934–52.
- Wallis JD, Anderson KC, Miller EK. 2001. Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–6.
- Wang M, Vijayraghavan S, Goldman-Rakic PS. 2004. Selective D2 receptor actions on the functional circuitry of working memory. *Science* 303:853–6.
- Wang XJ. 1999. Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J Neurosci* 19:587–603.
- Wang XJ. 2001. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci* 24:455–63.
- Wang XJ, Tegner J, Constantinidis C, Goldman-Rakic PS. 2004. Division of labor among distinct inhibitory neurons in a cortical microcircuit of working memory. *Proc Natl Acad Sci U S A* 101:1368–73.
- White IM, Wise SP. 1999. Rule-dependent neuronal activity in the prefrontal cortex. *Exp Brain Res* 126:315–35.
- Williams GV, Goldman-Rakic PS. 1995. Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. *Nature* 376:572–5.
- Wilson FA, O'Scalaidhe SP, Goldman-Rakic PS. 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260:1955–8.
- Wilson FA, O'Scalaidhe SP, Goldman-Rakic PS. 1994. Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proc Natl Acad Sci U S A* 91:4009–13.
- Wood SJ, Pantelis C, Proffitt T, Phillips LJ, Stuart GW, Buchanan JA, and others. 2003. Spatial working memory ability is a marker of risk-for-psychosis. *Psychol Med* 33:1239–47.

The Bayesian brain: the role of uncertainty in neural coding and computation

David C. Knill and Alexandre Pouget

Center for Visual Science and the Department of Brain and Cognitive Science, University of Rochester, NY 14627, USA

To use sensory information efficiently to make judgments and guide action in the world, the brain must represent and use information about uncertainty in its computations for perception and action. Bayesian methods have proven successful in building computational theories for perception and sensorimotor control, and psychophysics is providing a growing body of evidence that human perceptual computations are 'Bayes' optimal'. This leads to the 'Bayesian coding hypothesis': that the brain represents sensory information probabilistically, in the form of probability distributions. Several computational schemes have recently been proposed for how this might be achieved in populations of neurons. Neurophysiological data on the hypothesis, however, is almost non-existent. A major challenge for neuroscientists is to test these ideas experimentally, and so determine whether and how neurons code information about sensory uncertainty.

Humans and other animals operate in a world of sensory uncertainty. Although introspection tells us that perception is deterministic and certain, many factors contribute to limiting the reliability of sensory information about the world – the mapping of 3D objects into a 2D image, neural noise introduced in early stages of sensory coding, and structural constraints on neural representations and computations (e.g. the density of receptors in the retina). Our brains must effectively deal with the resulting uncertainty to generate perceptual representations of the world and to guide our actions. This leads naturally to the idea that perception is a process of unconscious, probabilistic inference [1,2]. Aided by developments in statistics and artificial intelligence, researchers have begun to apply the concepts of probability theory rigorously to problems in biological perception and action [3–20]. One striking observation from this work is the myriad ways in which human observers behave as optimal Bayesian observers. This observation, along with the behavioral and computational work on which it is based, has fundamental implications for neuroscience, particularly in how we conceive of neural computations and the nature of neural representations of perceptual and motor variables.

Bayesian inference and the Bayesian coding hypothesis

The fundamental concept behind the Bayesian approach to perceptual computations is that the information provided by a set of sensory data about the world is represented by a conditional probability density function over the set of unknown variables – the posterior density function. A Bayesian perceptual system, therefore, would represent the perceived depth of an object, for example, not as a single number Z but as a conditional probability density function $p(Z/I)$, where I is the available image information (e.g. stereo disparities). Loosely speaking, $p(Z/I)$ would specify the relative probability that the object is at different depths Z , given the available sensory information.

More generally, the component computations that underlay Bayesian inferences [that give rise to $p(Z/I)$] are ideally performed on representations of conditional probability density functions rather than on unitary estimates of parameter values. Loosely speaking, a Bayes' optimal system maintains, at each stage of local computation, a representation of all possible values of the parameters being computed along with associated probabilities. This allows the system to integrate information efficiently over space and time, to integrate information from different sensory cues and sensory modalities, and to propagate information from one stage of processing to another without committing too early to particular interpretations. Bayesian statisticians refer to the idea of representing and propagating information in the form of conditional density functions as belief propagation, and this approach has been highly successful in designing effective artificial vision systems [21–23].

To illustrate the basic structure of Bayesian computations, consider the problem of integrating multiple sensory cues about some property of a scene. Figure 1 illustrates the Bayesian formulation of one such problem – estimating the position of an object X from visual and auditory cues V and A . The goal of an optimal, Bayesian observer would be to compute the conditional density function $p(X/V,A)$. Using Bayes' rule, this is given by

$$P(X/V,A) = p(V,A/X)p(X)/p(V,A) \quad (\text{Equation 1})$$

where $p(V,A/X)$ specifies the relative likelihood of sensing the given data for different values of X and $p(X)$ is the prior probability of different values of X . Because the noise

Corresponding author: David C. Knill (knill@cvs.rochester.edu).

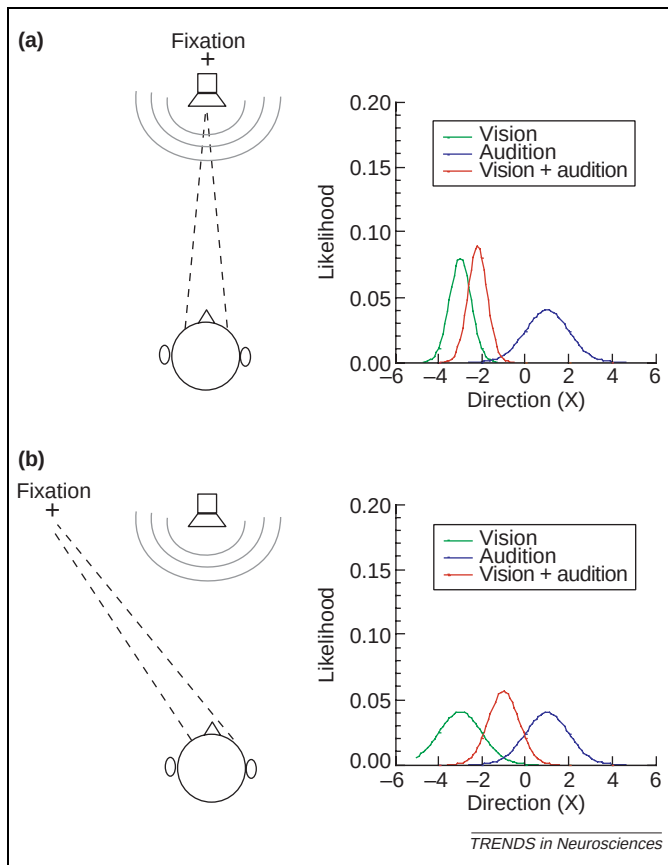


Figure 1. Two examples in which auditory and visual cues provide ‘conflicting’ information about the direction of a target. The conflict is apparent in the difference in means of the likelihood functions associated with each cue, although the functions overlap. Such conflicts are always present, owing to noise in the sensory systems. To integrate visual and auditory information optimally, a multimodal area must take into account the uncertainty associated with each cue. (a) When the vision cue is most reliable, the peak of the posterior distribution is shifted toward the direction suggested by the vision cue. (b) When the reliabilities of the cues are more similar, for example when the stimulus is in the far periphery, the peak is shifted toward the direction suggested by the auditory cue. When both likelihood functions are Gaussian, the most likely direction of the target is given by a weighted sum of the most likely directions (μ) given the visual (V) and auditory (A) cues individually: $\mu_{V,A} = w_V \mu_V + w_A \mu_A$. The weights (w) are inversely proportional to the variances of the likelihood functions.

sources in auditory and visual mechanisms are statistically independent, we can decompose the likelihood function into the product of likelihood functions associated with the visual and auditory cues, respectively:

$$P(V,A/X) = p(V/X)p(A/X) \quad (\text{Equation 2})$$

$p(V/X)$ and $p(A/X)$ fully represent the information provided by the visual and auditory data about the position of the target. The posterior density function is therefore proportional to the product of three functions: the likelihood functions associated with each cue and the prior density function representing the relative probability of the target being at any given position. An optimal estimator could pick the peak of the posterior density function, the mean of the function or any of several other choices, depending on the cost associated with making different types of errors [24].

For our purposes, the point of the example is that an optimal integrator must take into account the relative uncertainty of each cue when deriving an integrated estimate. When one cue is less certain than another, the integrated estimate should be biased toward the more

reliable cue. Assuming that a system can accurately compute and represent likelihood functions, the calculation embodied in equations 1 and 2 implicitly enforces this behavior (Figure 1). Although other estimation schemes can show the same performance as an optimal Bayesian observer (e.g. a weighted sum of estimates independently derived from each cue), computing with likelihood functions provides the most direct means available to account ‘automatically’ for the large range of differences in cue uncertainty that an observer is likely to face.

This is the basic premise on which Bayesian theories of cortical processing will succeed or fail – that the brain represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions. We will refer to this as the ‘Bayesian coding hypothesis’. The opposing view is that neural representations are deterministic and discrete, which might be intuitive but also misleading. This intuition might be due to the apparent ‘oneness’ of our perceptual world and the need to ‘collapse’ perceptual representations into discrete actions, such as decisions or motor behaviors. The principle data on the Bayesian coding hypothesis are behavioral results showing the many different ways in which humans perform as Bayesian observers.

Are human observers Bayes’ optimal?

What does it mean to say that an observer is ‘Bayes’ optimal’? Humans are clearly not optimal in the sense that they achieve the level of performance afforded by the uncertainty in the physical stimulus. Absolute efficiencies (a measure of performance relative to a Bayes’ optimal observer) for performing high-level perceptual tasks are generally low and vary widely across tasks. In some cases, this inefficiency is entirely due to uncertainty in the coding of sensory primitives that serve as inputs to perceptual computations [6]; in others, it is due to a combination of sensory, perceptual and cognitive factors [25]. The real test of the Bayesian coding hypothesis is in whether the neural computations that result in perceptual judgments or motor behavior take into account the uncertainty in the information available at each stage of processing. Psychophysical work in several areas suggests that this is the case.

Cue integration

Perhaps the most persuasive evidence for the Bayesian coding hypothesis comes from work on sensory cue integration. When the uncertainty associated with each of a set of cues is approximated by a Gaussian likelihood function, the average estimate derived from an optimal Bayesian integrator is a weighted average of the average estimates that would be derived from each cue alone (Figure 1). The reliability of different cues changes as a function of many scene and viewing parameters (e.g. the reliability of stereo disparity decreases with viewing distance). When these parameters vary from trial to trial in a psychophysical experiment, an optimal Bayesian observer would appear to weight cues differently on different trials. Studies of human cue integration, both within modality (e.g. stereo and texture) [26–28] and

across modality (e.g. sight and touch or sight and sound) [29–32], consistently find cue weights that vary in the manner predicted by Bayesian theory. Although these results could be accounted for by a deterministic system that adjusts cue weights as a function of viewing parameters and stimulus properties that co-vary with cue uncertainty, representing and computing with probability distributions (Figure 1) is considerably more flexible and can accommodate novel stimulus changes that alter cue uncertainty.

Non-linear Bayesian estimation

One of the strongest computational arguments for representing density functions in intermediate perceptual computations is that they are often not Gaussian, so that simple linear mechanisms do not suffice to support optimal Bayesian calculations. Non-Gaussian likelihood functions arise even when the sensory noise is Gaussian as a result of the nonlinear mapping from sensory feature space to the parameter space being estimated. In these cases, computations on density functions (or likelihood functions) are necessary to achieve optimality [33].

Figure 2 shows a simple example in the context of cue integration. Changing the angle of a symmetric figure within its plane (its spin), keeping the 3D orientation of the plane itself fixed (imagine spinning a rectangle around a rod oriented perpendicular to the rectangle), changes the perceived 3D orientation of the plane, even when viewed in stereo [16]. These spin-dependent biases are well accounted for by a Bayesian model that optimally combines skew symmetry information (represented by a highly non-Gaussian likelihood function in Figure 2) with stereoscopic information about 3D surface orientation. The results would not be predicted by a deterministic scheme of weighting the estimates derived from each cue individually.

Perceptual biases and priors

Several recent studies on the role of prior models in perception provide strong evidence for Bayesian computations of the type envisioned here [15,18]. Weiss and Adelson's [17] recent work on motion provides an illustrative example. They propose a remarkably simple Bayesian model of motion perception in which likelihood functions derived from local, ambiguous motion measurements are multiplied together with a simple prior that biases interpretations to favor low speeds. The interplay between the likelihood functions and the prior density function leads to a complex pattern of directional biases that depends on contrast, edge orientation and other stimulus factors. The model predicts a surprisingly large range of previously unintuitive motion phenomena, suggesting that the brain could perform a similar computation.

Uncertainty and the control of action

The previous examples were all based on the performance of subjects in perceptual tasks. Sensory information, however, primarily serves the function of guiding action in the world. Researchers have recently begun developing techniques for coupling optimal Bayesian estimators to

control systems to maximize performance in motor tasks [34]. In much the same way that sensory uncertainty determines the optimal weighting scheme for combining sensory cues for perceptual judgments, sensory and motor uncertainties determine how sensory signals should be used to plan and control movements.

Consider the problem of using sensory feedback from the hand to guide online corrections of hand movements. Because of noise in the motor system and in initial sensory estimates of hand and target position, the movements of an individual are never perfect. Visual feedback from the moving hand should, in theory, be used to make small adjustments to movement trajectories online. How much an observer should trust the visual feedback, however, depends on how reliable it is. For example, when the hand is in the periphery of the visual field, visual estimates of position will necessarily be worse than when it is in the foveal area of the field (and near the target). Similarly, the error in motion signals from the hand scales with velocity, so that motion signals are least reliable around the point of peak velocity. Recent psychophysical studies have shown that humans use continuous feedback from the hand to control pointing movements, but that the relative contributions of different signals (e.g. position and velocity) depend on the expected sensory noise associated with those signals. Moreover, when noise is artificially added to visual feedback about the position of the hand, subjects optimally adjust the degree to which they rely on the feedback to make corrections [14,35,36].

Likewise, how one plans movements should depend on the intrinsic variability in the motor output and the costs associated with various errors. Recent behavioral tests have confirmed that motor plans take into account the uncertainty in motor outputs: ballistic movement trajectories effectively minimize the error in pointing movements, given the signal-dependent properties of motor noise [8], and when costs and gains for different aim points are independently varied, subjects adjust their aim points for fast pointing movements to maximize the expected gain [19,20]. Although not definitive, these results suggest that the brain uses knowledge of the uncertainty in the sensory input and the motor output for visuomotor control.

Neural representations of uncertainty

The notion that neural computations take into account the uncertainty of the sensory and motor variables raises two important questions: (i) how do neurons, or rather populations of neurons, represent uncertainty, and (ii) what is the neural basis of statistical inferences? Several schemes have been proposed over the past few years, which we now briefly review.

Binary variables

The simplest schemes apply to binary variables – that is, variables that can take only two states. This situation arises when subjects are asked to decide between two possibilities, such as whether an object is moving up or down. In this case, the uncertainty of the variable can be encoded in two ways. First, one can use two populations of neurons: one in which neurons respond proportionally to the probability of upward motion, and one in which they

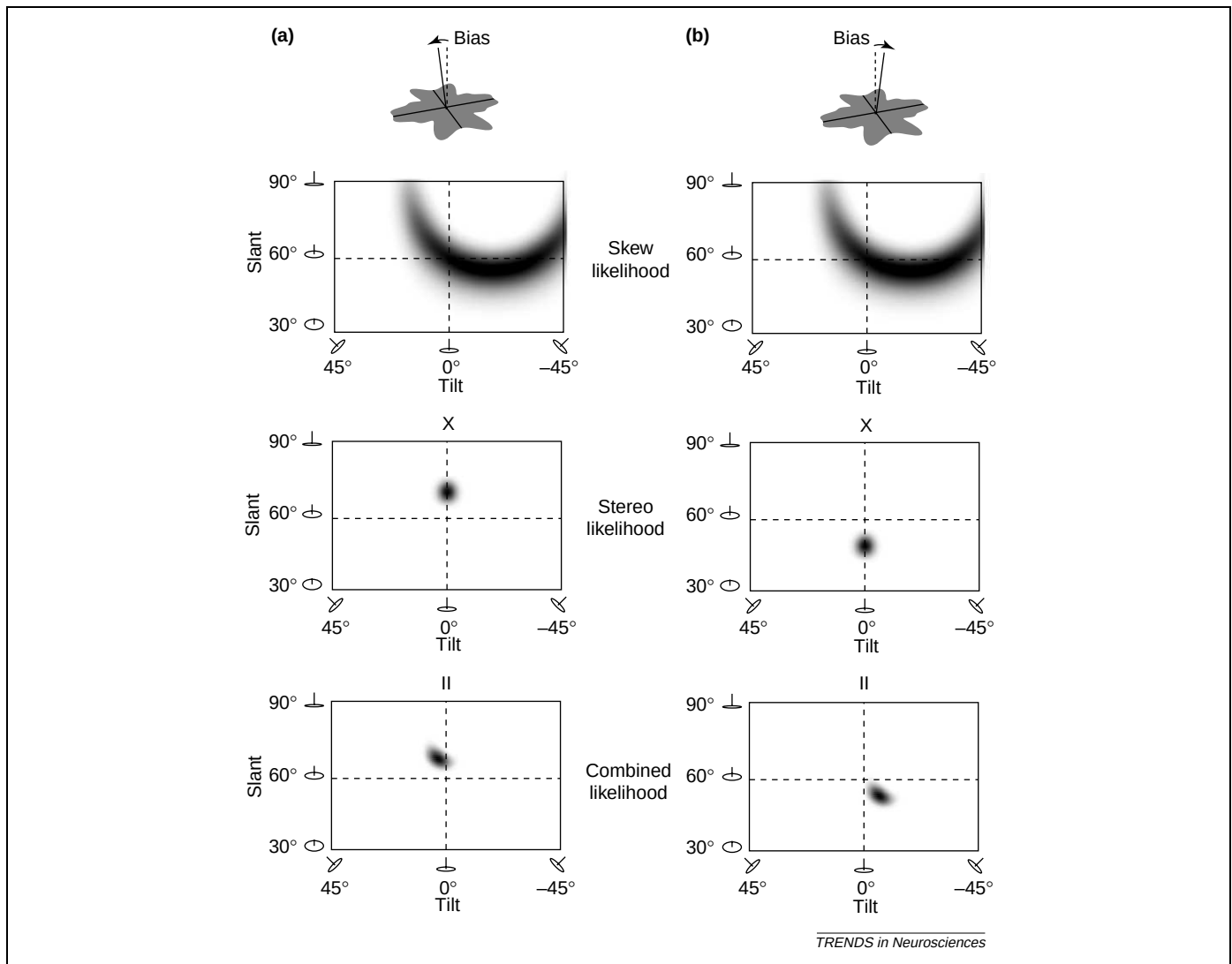


Figure 2. Skew symmetrical figures appear as figures slanted in depth because the brain assumes that they are projected from bilaterally symmetrical figures in the world. The information provided by skew symmetry is given by the angle between the projected symmetry axes of a figure, shown at the top of each panel as solid lines superimposed on the figure. Assuming that visual measurements of the orientations of these angles in the image are corrupted by Gaussian noise, one can compute a likelihood function for 3D surface orientation from skew. The result, as shown on the top two graphs, is highly non-Gaussian (darker points indicate higher likelihood values). The shape of the likelihood function is highly dependent on the spin of the figure around its 3D surface normal. When combined with stereoscopic information from binocular disparities (middle), an optimal estimator multiplies the likelihood functions associated with skew and stereo, as indicated by the 'X' between the top and middle panels, to produce a posterior distribution for surface orientation (bottom) given both cues (assuming the prior on surface orientation is flat). This leads to the prediction that perceptual biases will depend on the spin of the figure. It also leads to the somewhat counterintuitive prediction illustrated here that changing the slant suggested by stereo disparities should change the perceived tilt of symmetric figures. (a) When binocular disparities suggest a slant greater than that from which a figure was projected, the posterior distribution is shifted away from the orientation suggested by the disparities in both slant and tilt, creating a biased percept of the tilt of the figure. (b) The same figure, when binocular disparities suggest a smaller slant, gives rise to a tilt bias in the opposite direction. This is exactly the pattern of behavior shown by subjects.

respond to the probability of downward motion. An even simpler scheme involves only one population in which neurons respond proportionally to the ratio of upward and downward motion, equivalent to a quantity known as a likelihood ratio. Single-cell recordings in the lateral intraparietal (LIP) area provide evidence for both schemes. When monkeys are trained to perform one of two possible saccades, a subset of LIP neurons respond proportionally to the probability that the saccade ends in their receptive field [37]. However, when monkeys are trained to distinguish between two possible directions of motion, some LIP neurons integrate information over time in a way consistent with the computation of a likelihood ratio [38]. Similar ideas have also been suggested to interpret neuronal responses in the superior colliculus [39].

Convolution codes and variations

When direction of motion is not confined to two choices, but can take any value around a circle, the previous schemes no longer work. Rather, the brain needs an encoding mechanism that can deal with continuous variables. As described previously, the most natural way to represent uncertainty in a continuous variable is to represent the probability density function of the variable (Figure 3a). Any probability density function, $p(x)$, can be represented by its value at a few sample points along the x axis. The more samples are used, the better the approximation. This is the idea behind convolution codes, except that convolution codes avoid gaps between the samples by filtering $p(x)$ with a Gaussian kernel before sampling [40,41]. Each neuron simply

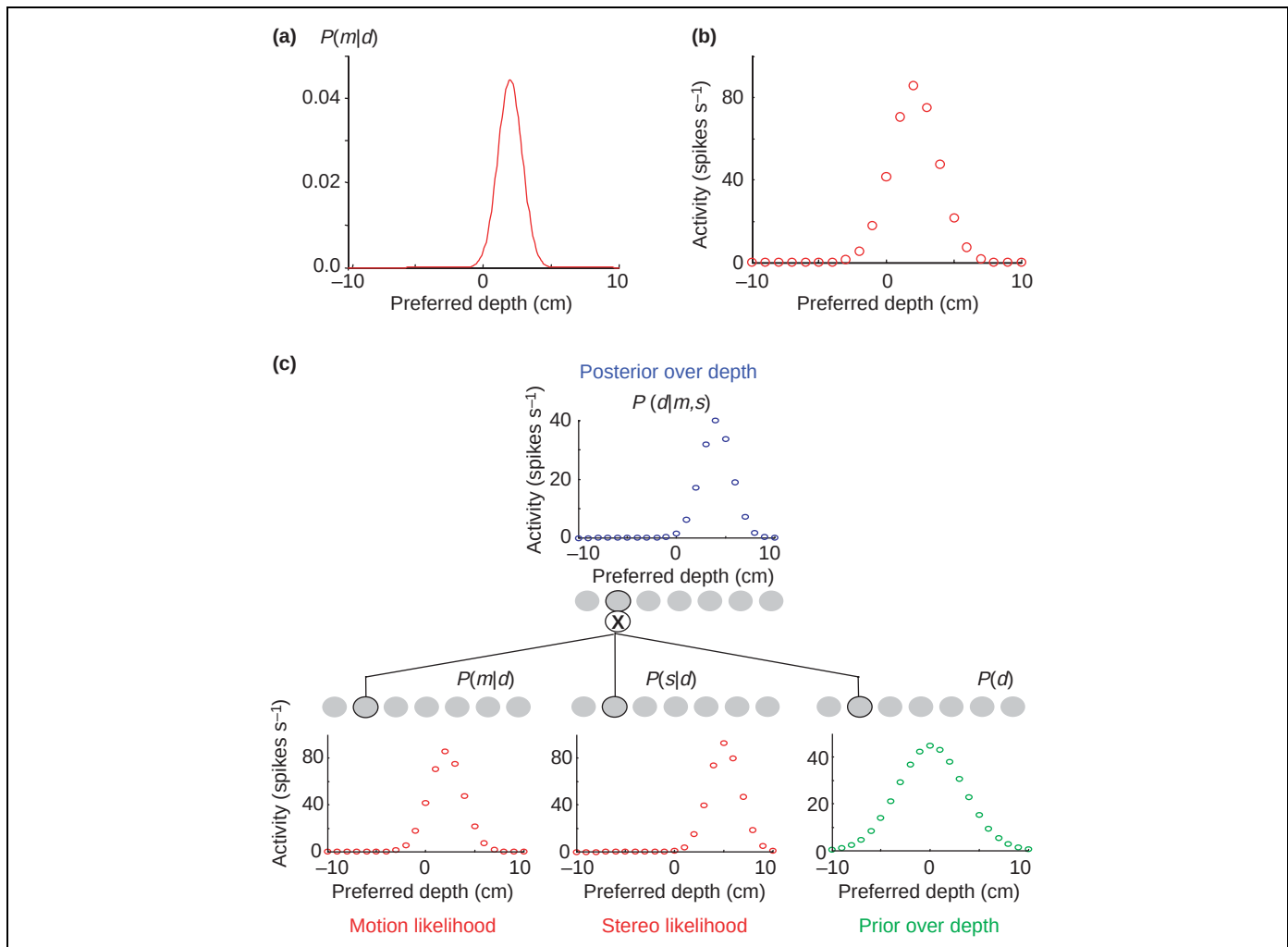


Figure 3. Inferences with convolution codes. (a) A hypothetical probability density function over motion given the depth of an object, $P(m|d)$. (b) A convolution code for the probability density function in (a). Each dot corresponds to the activity of one neuron plotted at its preferred depth. The code is basically a set of samples of the probability density function filtered, or convolved, by a Gaussian kernel. The width of this pattern of activity is related to the width of the function and, therefore, to the uncertainty of the encoded variable. (c) Given two likelihood functions $P(m|d)$ and $P(s|d)$, and one prior distribution $P(d)$, the posterior distribution over depth $P(d|m,s)$, is obtained by taking a neuron-by-neuron product of the three distributions (neurons are indicated by shaded circles, ranked by their preferred depth). This scheme works only if the Gaussian kernel used in the convolution code is a Dirac function, but it can easily be adapted to Gaussian kernels of any width [43]. Adapted from Ref. [46] © (2003) Annual Reviews (www.annualreviews.org).

computes the dot product between its Gaussian tuning curve and the probability density function. Assuming that the tuning curves of the neurons are just translated copies of the same Gaussian profile, the resulting population pattern of activity looks like the original probability density function filtered by the Gaussian tuning curve.

Inferences with convolution codes are straightforward. For instance, given motion and stereo disparity cues for depth, one would like to compute the posterior density function for depth conditioned on motion and stereo information by multiplying the likelihood functions for the observed motion of an object given its depth $p(m|d)$ with the observed pattern of horizontal disparities given its depth $p(s|d)$ and a prior distribution over depth. If neurons represent samples of the likelihood functions and the prior density function, a simple point-by-point product operation between the two representations (Figure 3c) is equivalent to multiplying the functions themselves [30,42]. In reality, this scheme works only when the tuning curves of the neurons are Dirac functions (infinitely narrow Gaussian tuning curves) but it can be easily extended to

wider tuning curves [42,43]. This scheme has also been applied to static and time-varying variables [40,44].

Another possibility for representing a probability density function with a population code is to encode the log of the probability density function, instead of the function itself [45]. With this scheme, the point-by-point product required for Bayesian inference with convolution codes (Figure 3c) is replaced by a simple point-by-point sum, because logs turn products into sums. Using sums is appealing because of its simplicity and because it is consistent with the way LIP neurons integrate (i.e. sum) evidence over time [38].

Gain encoding

An alternative to the convolution code is to use what we call a gain-encoding scheme [46,47]. This scheme takes advantage of the near-Poisson nature of neural noise [48] to code the mean and variance of a density function simultaneously. To understand how the scheme works, consider the example of orientation selectivity. The primary visual cortex contains neurons with bell-shaped

tuning curves for orientation [49]. If we rank the neurons by their preferred orientations, the population response to a trial of particular orientation θ_0 takes the form of a hill of activity (Figure 4b). On any given trial, the shape of the hill is corrupted by near-Poisson noise. To decode such noisy population codes, one can use a Bayesian decoder which returns the posterior distribution over θ given the hill of activity, $p(\theta|A)$ [50,51]. For independent Poisson noise, the posterior distribution is Gaussian, with its mean controlled mostly by the position of the peak of the hill and the variance inversely proportional to the gain of the hill [46]. This is because, for Poisson noise, the variance of the spike count is proportional to the gain. This implies that the signal-to-noise ratio – the ratio of the gain over the square root of the variance – grows with the square root of the gain. Therefore, a high gain entails a high signal-to-noise ratio, and a narrow posterior distribution. Consequently, the noisy hill of activity can be treated as a neural code for the posterior, with the position of the peak encoding the mean, and the amplitude (or gain) encoding the variance.

Deneve *et al.* [47] have designed a network architecture that uses gain-encoding to perform optimal Bayesian inferences. They applied their network to the problem of locating an object based on its sound and image. On each trial, the network is initialized with two noisy population codes for the position of an object in visual and auditory coordinates (Figure 5). It then performs two tasks. First, it remaps the visual input into auditory coordinates and vice versa, through a basis function layer. This is a prerequisite for combining these signals because the visual system encodes the location of the object in retinal coordinates whereas the auditory system uses head-centered coordinates. The basis function units act as the building blocks of the transformation: they compute Gaussian functions of the visual and auditory input that are combined to approximate both changes of coordinates, just as a set of cubes can be combined to approximate any three dimensional shape. Second, the network recovers the maximum likelihood estimate of the object position given the visual and auditory inputs (Figure 5). This computation is the result of a relaxation process that turns the noisy population codes into smooth hills of activity over time. For a particular value of the network parameters, these smooth hills of activity peak very close to the maximum likelihood estimate

of the position. The cues are integrated with weights proportional to their reliability because noisy hills with high gain – corresponding to more reliable cues – provide a stronger initial push and, as a result, have a stronger influence of the final state of the network [47,52].

Although originally applied to object localization, this architecture can be generalized to any cue integration problem. In particular, this approach can be used to account for the performance of human observers in the experiments on cue integration [26–32]. The model can also be extended to time varying problems, such as estimating the position of a moving arm [4].

The gain-encoding model suggests a particularly intriguing role for Poisson variability. At first sight, it would appear that this variability is highly detrimental and severely limits the accuracy with which cortical circuits perform computations. The gain-encoding idea suggests that Poisson noise might in fact be very beneficial: it allows population codes to represent the mean as well as the variance of the encoded variables, the latter being crucial for Bayesian inferences.

It is important to emphasize that the different encoding schemes we have reviewed are not mutually exclusive. Uncertainties can take many forms; for example, the uncertainty due to photon noise in the retina has little to do with the ambiguity due to the aperture problem in motion processing. It is therefore possible that the brain uses multiple encoding schemes. Ultimately, which schemes are used in the brain can be answered only empirically. It is our hope that the accumulation of behavioral data showing that neural computation is akin to a Bayesian inference, and the development of several models of Bayesian inference in neural networks, will compel neurophysiologists to design experiments to test the predictions of these models.

Discussion

We have described psychophysical evidence that shows human observers to behave in a variety of ways like optimal Bayesian observers. The most compelling features of these data in regard to the Bayesian coding hypothesis are: (i) that subjects implicitly ‘adjust’ cue weights in a Bayes’ optimal way based on stimulus and viewing parameters; (ii) that perceptual and motor behavior reflect a system that takes into account the uncertainty of both sensory and motor signals; (iii) that humans behave

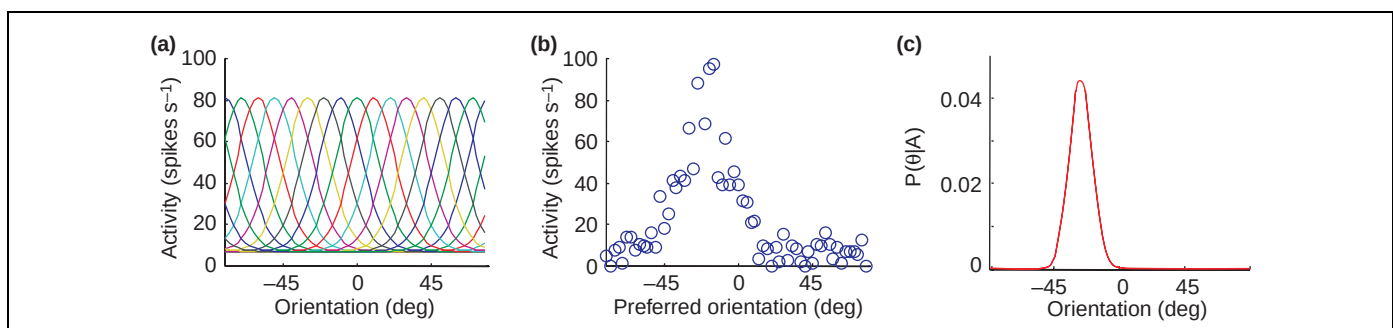


Figure 4. Inferences with gain encoding. (a) Idealized Gaussian tuning curves to orientation for 16 cells in primary visual cortex. (b) Response of 64 cells with Gaussian tuning curves similar to the ones shown in (a), in response to an orientation of -20° . The cells have been ranked according to their preferred orientation and the responses have been corrupted by independent Poisson noise, a good approximation of the noise observed *in vivo*. (c) The posterior distribution over orientation obtained from applying a Bayesian decoder to the noisy hills shown in (b). With independent Poisson noise, the peak of the distribution is given by the peak of the noisy hill, and the width of the distribution (i.e. the uncertainty) is inversely proportional the amplitude of the noisy hill. Adapted from Ref. [47].

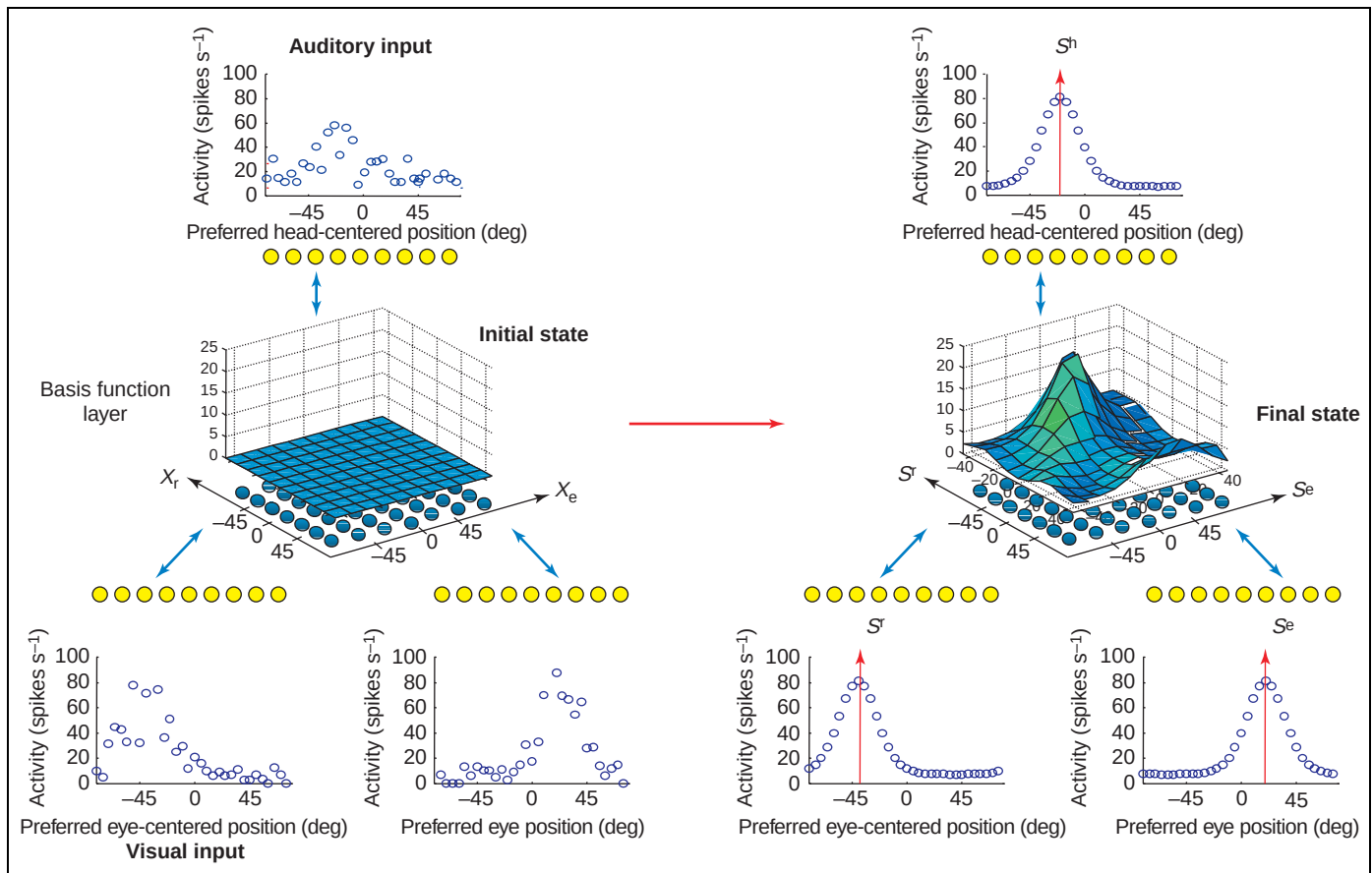


Figure 5. A basis function network for optimal cue integration between a visual and an auditory input. The network is tuned to perform two tasks simultaneously. Through the basis function layer, eye-centered visual inputs (X_e) are remapped into auditory coordinates (head-centered) and vice versa (this requires knowledge of the position of the eyes, which is encoded in a third population code, X_o). Over time, the network settles onto three smooth hills of activity, peaking in the final state at the location of the maximum likelihood estimates of the eye-centered position (S^e) and the head-centered position (S^h) of the target and at the position of the eyes in the head (S^o). This solution takes into account the respective reliability of each of the cues, as expected for a maximum likelihood estimate. In the case illustrated here, the auditory input is less reliable than the visual one, as indicated by the fact that the auditory hill has a reduced gain (top left). Adapted from Ref. [47].

near-optimally even when the sensory information is characterized by highly non-Gaussian density functions, leading to complex patterns of predicted behavior; and (iv) that relatively simple Bayesian models can account for otherwise complex patterns of subjective, perceptual biases, even when the prior density functions built into the models do not explicitly code the biases. We argue that these data strongly suggest that the brain codes even complex patterns of sensory and motor uncertainty in its internal representations and computations. Two challenges for further research emerge from this review.

First, although a growing body of psychophysical work is being devoted to exploring the ways in which humans are optimal observers and actors, an equally important challenge for future work is to find the ways in which human observers are not optimal. Owing to the complexity of the tasks, unconstrained Bayesian inference is not a viable solution for computation in the brain. Recent work on statistical learning, for example, has elucidated strong limits on the types of statistical regularities that sensory systems automatically detect [53–55].

Second, neuroscientists must begin to test theories of how uncertainty could be represented in populations of neurons. We have described several neural coding strategies that might be used to encode probability density

functions or their statistical moments; however, the neurophysiological evidence for these schemes is weak. This is not because existing data conflict with the strategies but, rather, because little work has been done to test them. Pursuing this challenge will require further development and application of advanced multi-electrode recording techniques. As these techniques mature, we hope that neuroscientists will take up the challenge to submit the Bayesian coding hypothesis to rigorous falsification tests.

Finally, we acknowledge that the examples presented here are much simpler than many of the perceptual estimation problems faced by the brain. Most notable among the complexities of these problems is their high dimensionality (e.g. contour completion and flow field estimation). Until recently, the problem of efficiently representing and computing with probability density functions in high-dimensional spaces has been a barrier to developing efficient Bayesian computer vision algorithms. The recent development of graphical models [22] and particle filtering techniques [56,57] has shown the most promise for implementing efficient Bayesian algorithms for high-dimensional estimation problems. How these might be implemented by the brain is a major challenge for computational neuroscientists [58].

References

- 1 Mach, E. (1980) *Contributions to the Analysis of the Sensations* (C. M. Williams, Trans.), Open Court Publishing Co., Chicago, IL USA
- 2 Helmholtz, H. (1925) *Physiological Optics, Vol. III: The perceptions of Vision* (J. P. Southall, Trans.), Optical Society of America, Rochester, NY USA
- 3 Liu, Z. *et al.* (1995) Object classification for human and ideal observers. *Vision Res.* 35, 549–568
- 4 Wolpert, D.M. *et al.* (1995) An internal model for sensorimotor integration. *Science* 269, 1880–1882
- 5 Crowell, J.A. and Banks, M.S. (1996) Ideal observer for heading judgments. *Vision Res.* 36, 471–490
- 6 Eagle, R.A. and Blake, A. (1995) 2-Dimensional constraints on 3-dimensional structure-from-motion tasks. *Vision Res.* 35, 2927–2941
- 7 Knill, D.C. (1998) Discrimination of planar surface slant from texture: human and ideal observers compared. *Vision Res.* 38, 1683–1711
- 8 Harris, C.M. and Wolpert, D.M. (1998) Signal-dependent noise determines motor planning. *Nature* 394, 780–784
- 9 van Beers, R.J. *et al.* (2001) Sensorimotor integration compensates for visual localization errors during smooth pursuit eye movements. *J. Neurophysiol.* 85, 1914–1922
- 10 van Beers, R.J. *et al.* (2002) Role of uncertainty in sensorimotor control. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1137–1145
- 11 Shimozaki, S.S. *et al.* (2003) An ideal observer with channels versus feature-independent processing of spatial frequency and orientation in visual search performance. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 2197–2215
- 12 Beutter, B.R. *et al.* (2003) Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1341–1355
- 13 Murray, R.F. *et al.* (2003) Saccadic and perceptual performance in visual search tasks. II. Letter discrimination. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1356–1370
- 14 Saunders, J.A. and Knill, D.C. (2004) Visual feedback control of hand movements. *J. Neurosci.* 24, 3223–3234
- 15 Mamassian, P. and Landy, M.S. (2001) Interaction of visual prior constraints. *Vision Res.* 41, 2653–2668
- 16 Saunders, J.A. and Knill, D.C. (2001) Perception of 3D surface orientation from skew symmetry. *Vision Res.* 41, 3163–3183
- 17 Weiss, Y. *et al.* (2002) Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604
- 18 van Ee, R. *et al.* (2003) Bayesian modeling of cue interaction: bistability in stereoscopic slant perception. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1398–1406
- 19 Trommershauser, J. *et al.* (2003) Statistical decision theory and trade-offs in the control of motor response. *Spat. Vis.* 16, 255–275
- 20 Trommershauser, J. *et al.* (2003) Statistical decision theory and the selection of rapid, goal-directed movements. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1419–1433
- 21 Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann Publishers, San Mateo CA, USA
- 22 Weiss, Y. and Freeman, W.T. (2001) Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Comput.* 13, 2173–2200
- 23 Freeman, W.T. *et al.* (2000) Learning low-level vision. *Int. J. Comput. Vis.* 40, 25–47
- 24 Geisler, W.S. and Diehl, R.L. (2003) A Bayesian approach to the evolution of perceptual and cognitive systems. *Cogn. Sci.* 27, 379–402
- 25 Geisler, W.S. (1989) Sequential ideal-observer analysis of visual discriminations. *Psychol. Rev.* 96, 267–314
- 26 Jacobs, R.A. (1999) Optimal integration of texture and motion cues to depth. *Vision Res.* 39, 3621–3629
- 27 Knill, D.C. and Saunders, J.A. (2003) Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Res.* 43, 2539–2558
- 28 Hillis, J.M. *et al.* Slant from texture and disparity cues: optional cue combination. *J. Vis.* (in press)
- 29 van Beers, R.J. *et al.* (1999) Integration of proprioceptive and visual position-information: an experimentally supported model. *J. Neurophysiol.* 81, 1355–1364
- 30 Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
- 31 Battaglia, P.W. *et al.* (2003) Bayesian integration of visual and auditory signals for spatial localization. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1391–1397
- 32 Alais, D. and Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262
- 33 Knill, D.C. (2003) Mixture models and the probabilistic structure of depth cues. *Vision Res.* 43, 831–854
- 34 Todorov, E. and Jordan, M.I. (2002) Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5, 1226–1235
- 35 Saunders, J.A. and Knill, D.C. (2003) Humans use continuous visual feedback from the hand to control fast reaching movements. *Exp. Brain Res.* 152, 341–352
- 36 Kording, K.P. and Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning. *Nature* 427, 244–247
- 37 Platt, M.L. and Glimcher, P.W. (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238
- 38 Gold, J.I. and Shadlen, M.N. (2001) Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16
- 39 Anastasio, T.J. *et al.* (2000) Using Bayes' rule to model multisensory enhancement in the superior colliculus. *Neural Comput.* 12, 1165–1187
- 40 Anderson, C. (1994) Neurobiological computational systems. In *Computational Intelligence Imitating Life*, pp. 213–222, IEEE Press, New York NY, USA
- 41 Zemel, R. *et al.* (1998) Probabilistic interpretation of population code. *Neural Comput.* 10, 403–430
- 42 Zemel, R.S. and Dayan, P. (1997) Combining probabilistic population codes. In *JCAI-97: 15th International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Francisco CA, USA
- 43 Barber, M.J. *et al.* (2003) Neural representation of probabilistic information. *Neural Comput.* 15, 1843–1864
- 44 Eliasmith, C. and Anderson, C.H. (2003) *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*, MIT Press
- 45 Rao, R.P. (2004) Bayesian computation in recurrent neural circuits. *Neural Comput.* 16, 1–38
- 46 Pouget, A. *et al.* (2003) Inference and computation with population codes. *Annu. Rev. Neurosci.* 26, 381–410
- 47 Deneve, S. *et al.* (2001) Efficient computation and cue integration with noisy population codes. *Nat. Neurosci.* 4, 826–831
- 48 Tolhurst, D. *et al.* (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* 23, 775–785
- 49 Hubel, D. and Wiesel, T. (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* 160, 106–154
- 50 Sanger, T. (1996) Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.* 76, 2790–2793
- 51 Foldiak (1993) The 'ideal homunculus': statistical inference from neural population responses. In *Computation and Neural Systems* (Eeckman, F. and Bower, J. eds), pp. 55–60, Kluwer Academic Publishers
- 52 Latham, P.E. *et al.* (2003) Optimal computation with attractor networks. *J. Physiol. Paris* 97, 683–694
- 53 Fiser, J. and Aslin, R.N. (2002) Statistical learning of new visual feature combinations by infants. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15822–15826
- 54 Fiser, J. and Aslin, R.N. (2002) Statistical learning of higher-order temporal structure from visual shape sequences. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 458–467
- 55 Fiser, J. and Aslin, R.N. (2001) Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychol. Sci.* 12, 499–504
- 56 Blake, A. *et al.* (1998) Statistical models of visual shape and motion. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 356, 1283–1301
- 57 Isard, M. and Blake, A. (1998) Condensation: conditional density propagation for visual tracking. *Int. J. Comput. Vis.* 29, 5–28
- 58 Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448

The following resources related to this article are available online at www.sciencemag.org (this information is current as of July 18, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/318/5850/606>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/318/5850/606#related-content>

This article **cites 33 articles**, 6 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/318/5850/606#otherarticles>

This article has been **cited by** 11 article(s) on the ISI Web of Science.

This article has been **cited by** 3 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/318/5850/606#otherarticles>

This article appears in the following **subject collections**:

Neuroscience

<http://www.sciencemag.org/cgi/collection/neuroscience>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

12. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
13. W. Schultz, *Annu. Rev. Psychol.* **57**, 87 (2006).
14. J. I. Gold, M. N. Shadlen, *Trends Cognit. Sci.* **5**, 10 (2001).
15. A. R. Damasio, *Sci. Am.* **271**, 144 (1994).
16. A. Bechara, H. Damasio, D. Tranel, A. R. Damasio, *Science* **275**, 1293 (1997).
17. A. Bechara, H. Damasio, A. R. Damasio, *Cereb. Cortex* **10**, 295 (2000).
18. A. D. (Bud) Craig, *Trends Neurosci.* **26**, 303 (2003).
19. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, *Nature* **441**, 876 (2006).
20. B. De Martino, D. Kumaran, B. Seymour, R. J. Dolan, *Science* **313**, 684 (2006).
21. G. F. Koob, M. Le Moal, *Science* **278**, 52 (1997).
22. M. Bateson, *Proc. Nutr. Soc.* **61**, 509 (2002).
23. C. J. Pietras, T. D. Hackenberg, *J. Exp. Anal. Behav.* **76**, 1 (2001).
24. G. Dom, B. Sabbe, W. Hulstijn, B. W. van den Brink, *Br. J. Psychiatry* **187**, 209 (2005).
25. G. J. Madden, W. K. Bickel, E. A. Jacobs, *Exp. Clin. Psychopharmacol.* **7**, 284 (1999).
26. S. F. Coffey, G. D. Gudleski, M. E. Saladin, K. T. Brady, *Exp. Clin. Psychopharmacol.* **11**, 18 (2003).
27. N. M. Petry, T. Casarella, *Drug Alcohol Depend.* **56**, 25 (1999).
28. N. M. Petry, *Psychopharmacology (Berlin)* **162**, 425 (2002).
29. G. Ainslie, in *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person* (Cambridge Univ. Press, New York, 1992), pp. 56–95.
30. K. N. Kirby, N. M. Petry, W. K. Bickel, *J. Exp. Psychol. Gen.* **128**, 78 (1999).
31. S. H. Kollins, *Addict. Behav.* **28**, 1167 (2003).
32. S. Grant, C. Contoreggi, E. D. London, *Neuropsychologia* **38**, 1180 (2000).
33. R. Gonzalez, A. Bechara, E. M. Martin, *J. Clin. Exp. Neuropsychol.* **29**, 155 (2007).
34. A. Verdejo-Garcia, R. Vilar-Lopez, M. Perez-Garcia, K. Podell, E. Goldberg, *J. Int. Neuropsychol. Soc.* **12**, 90 (2006).
35. A. Verdejo-Garcia, C. Rivas-Perez, R. Vilar-Lopez, M. Perez-Garcia, *Drug Alcohol Depend.* **86**, 139 (2007).
36. B. B. Quednow et al., *Psychopharmacology (Berlin)* **189**, 517 (2007).
37. G. Dom, B. De Wilde, W. Hulstijn, B. W. van den Brink, B. Sabbe, *Alcohol. Clin. Exp. Res.* **30**, 1670 (2006).
38. R. Pirastu et al., *Drug Alcohol Depend.* **83**, 163 (2006).
39. A. Bechara et al., *Neuropsychologia* **39**, 376 (2001).
40. R. D. Rogers et al., *Neuropsychopharmacology* **20**, 322 (1999).
41. M. P. Paulus, N. Hozack, L. Frank, G. G. Brown, M. A. Schuckit, *Biol. Psychiatry* **53**, 65 (2003).
42. J. Monterosso, R. Ehrman, K. L. Napier, C. P. O'Brien, A. R. Childress, *Addiction* **96**, 1825 (2001).
43. J. R. Monterosso et al., *Hum. Brain Mapp.* **28**, 383 (2007).
44. M. P. Paulus et al., *Neuropsychopharmacology* **26**, 53 (2002).
45. K. I. Bolla et al., *Neuroimage* **19**, 1085 (2003).
46. A. Bechara, *Nat. Neurosci.* **8**, 1458 (2005).
47. J. McCartney, *Subst. Use Misuse* **32**, 2061 (1997).
48. D. R. Cherek, S. D. Lane, *Psychopharmacology (Berlin)* **157**, 221 (2001).
49. D. S. Leland, M. P. Paulus, *Drug Alcohol Depend.* **78**, 83 (2005).
50. D. S. Leland, E. Arce, J. S. Feinstein, M. P. Paulus, *Neuroimage* **33**, 725 (2006).
51. S. Rahman et al., *Neuropsychopharmacology* **31**, 651 (2006).
52. C. N. Ortner, T. K. MacDonald, M. C. Olmstead, *Alcohol Alcohol.* **38**, 151 (2003).
53. J. B. Richards, L. Zhang, S. H. Mitchell, H. de Wit, *J. Exp. Anal. Behav.* **71**, 121 (1999).
54. H. de Wit, J. L. Enggasser, J. B. Richards, *Neuropsychopharmacology* **27**, 813 (2002).
55. F. X. Vollenweider, M. E. Liechti, M. P. Paulus, *J. Psychopharmacol.* **19**, 366 (2005).
56. B. Reynolds, J. B. Richards, M. Dassinger, H. de Wit, *Pharmacol. Biochem. Behav.* **79**, 17 (2004).
57. J. McDonald, L. Schleifer, J. B. Richards, H. de Wit, *Neuropsychopharmacology* **28**, 1356 (2003).
58. S. D. Lane, D. R. Cherek, O. V. Tcheremissine, L. M. Lieving, C. J. Pietras, *Neuropsychopharmacology* **30**, 800 (2005).
59. S. D. Lane, O. V. Tcheremissine, L. M. Lieving, S. Nouvion, D. R. Cherek, *Psychopharmacology (Berlin)* **181**, 364 (2005).
60. J. Epstein et al., *Am. J. Psychiatry* **163**, 1784 (2006).
61. E. E. Forbes et al., *J. Child Psychol. Psychiatry* **47**, 1031 (2006).
62. A. Must et al., *J. Affect. Disord.* **90**, 209 (2006).
63. T. Christodoulou, M. Lewis, G. B. Ploubidis, S. Frangou, *Eur. Psychiatry* **21**, 270 (2006).
64. F. C. Murphy et al., *Psychol. Med.* **31**, 679 (2001).
65. A. Minassian, M. P. Paulus, W. Perry, *J. Affect. Disord.* **82**, 203 (2004).
66. R. Ladouceur, P. Gosselin, M. J. Dugas, *Behav. Res. Ther.* **38**, 933 (2000).
67. I. Blanchette, A. Richards, *J. Exp. Psychol. Gen.* **132**, 294 (2003).
68. M. P. Paulus, J. S. Feinstein, A. Simmons, M. B. Stein, *Biol. Psychiatry* **55**, 1179 (2004).
69. A. L. Krain et al., *J. Child Psychol. Psychiatry* **47**, 1023 (2006).
70. S. Reiss, R. A. Peterson, D. M. Gursky, R. J. McNally, *Behav. Res. Ther.* **24**, 1 (1986).
71. M. P. Paulus, M. B. Stein, *Biol. Psychiatry* **60**, 383 (2006).
72. A. Simmons, S. C. Matthews, M. B. Stein, M. P. Paulus, *Neuroreport* **15**, 2261 (2004).
73. M. P. Paulus, C. Rogalsky, A. Simmons, J. S. Feinstein, M. B. Stein, *Neuroimage* **19**, 1439 (2003).
74. P. L. Remijnse et al., *Arch. Gen. Psychiatry* **63**, 1225 (2006).
75. N. S. Lawrence et al., *Neuropsychology* **20**, 409 (2006).
76. M. M. Nielen, D. J. Veltman, R. de Jong, G. Mulder, J. A. den Boer, *J. Affect. Disord.* **69**, 257 (2002).
77. K. D. Fitzgerald et al., *Biol. Psychiatry* **57**, 287 (2005).
78. P. S. Sachdev, G. S. Malhi, *Aust. N. Z. J. Psychiatry* **39**, 757 (2005).
79. O. H. Turnbull, C. E. Evans, K. Kemish, S. Park, C. H. Bowman, *Neuropsychology* **20**, 290 (2006).
80. S. B. Hutton et al., *Schizophr. Res.* **55**, 249 (2002).
81. M. P. Paulus, M. A. Geyer, D. L. Braff, *Am. J. Psychiatry* **153**, 714 (1996).
82. M. P. Paulus, M. A. Geyer, D. L. Braff, *Schizophr. Res.* **35**, 69 (1999).
83. M. P. Paulus, W. Perry, D. L. Braff, *Biol. Psychiatry* **46**, 662 (1999).
84. K. Ludewig, M. P. Paulus, F. X. Vollenweider, *Psychiatry Res.* **119**, 293 (2003).
85. M. P. Paulus, L. Frank, G. G. Brown, D. L. Braff, *Neuropsychopharmacology* **28**, 795 (2003).
86. S. Stroup et al., *Schizophr. Res.* **80**, 1 (2005).
87. A. D. Redish, *Science* **306**, 1944 (2004).
88. I would like to acknowledge the help of E. Arce, D. Leland, S. Matthews, M. Wittmann, and A. Simmons during the preparation of the manuscript. This research was supported by grants from the National Institute on Drug Abuse (R01DA016663 and R01DA018307) and a U.S. Department of Veterans Affairs merit grant.

10.1126/science.1142997

REVIEW

Decision Theory: What “Should” the Nervous System Do?

Konrad Körding

The purpose of our nervous system is to allow us to successfully interact with our environment. This normative idea is formalized by decision theory that defines which choices would be most beneficial. We live in an uncertain world, and each decision may have many possible outcomes; choosing the best decision is thus complicated. Bayesian decision theory formalizes these problems in the presence of uncertainty and often provides compact models that predict observed behavior. With its elegant formalization of the problems faced by the nervous system, it promises to become a major inspiration for studies in neuroscience.

Evolutionary psychology has found that many human behaviors can be well understood assuming adaptation of psychology to the past social environment of humans

[e.g., (1)]. Similarly, ethology, the study of animal behavior [e.g., (2)], has shown that many of the properties of the nervous system and the bodies of animals are remarkably well adapted to their eco-

logical niche. These disciplines have shown that, over the course of evolution, animals are often endowed with solutions to common problems that are close to optimal [(1), but see (3)]. Many studies in neuroscience analyze low-level processes. For example, researchers study how animals control their limbs, how they infer events in the world, and how they choose one of several possible rewards. Such processes may have remained conserved for very long periods of time. We can thus expect the solution used by the nervous system for such problems to be close to optimal.

Normative models formalize how the idea of adaptation predicts properties of the nervous system. These models assume that a process has

Department of Physical Medicine and Rehabilitation, Department of Physiology, and Department of Applied Mathematics, Institute of Neuroscience, Northwestern University and Rehabilitation Institute of Chicago, Room O-922, 345 East Superior Street, Chicago, IL 60611, USA. E-mail: kk@northwestern.edu

an objective (e.g., walk using least energy). Such an objective is typically formalized by a function (“utility”) that defines how good the solution is (e.g., the energy used). By combining the utility function with constraints (e.g., properties of muscles), it is possible to derive the best possible solution to the problem, usually using computer simulations. If the prediction of the model is matched by biology, it is concluded that, indeed, we have understood the purpose of a system. Normative models thus ask the “why?” question and formalize the ultimate purpose of a system. In this review, I focus on normative approaches to decision making; the nervous system often comes up with near-optimal decisions in a world characterized by uncertainty. To understand the nervous system, descriptive knowledge of how the nervous system works should be combined with normative knowledge of which problems it solves.

Decision Theory

The purpose of the central nervous system is to make decisions so that we can thrive by interacting successfully with our environment (Fig. 1A). When we play darts, we need to decide which position on the dartboard to aim at (Fig. 1B). The dartboard assigns a score to any possible dart position and thus defines the outcome. The objective of playing darts is to obtain as many points as possible. Within decision theory, such an objective is called a utility function [e.g., (4)]. A utility function $U(\text{outcome})$ measures how good or bad any possible decision outcome is. If dart players could choose where the dart will hit the board, they would choose the position that yields the most points and would thus maximize utility.

Although we can freely make decisions, we cannot directly choose the decision outcomes. If we always aim for the same position a , say the center of the bull’s eye, and throw many darts, we will produce a distribution of dart positions, x , on the dart board (Fig. 1C, inset). Within decision theory, this probability distribution is denoted $p(\text{outcome} = x | \text{decision} = a)$. If we aim at the position on the board that gives the highest score, we may instead hit a neighboring area of the dartboard and receive a low score. Depending on the position we aim at, different scores become more or less likely. This is a special case of a general problem in decision theory: Outcomes depend on decisions in a probabilistic way.

To derive the most beneficial decision, it is necessary to combine the utility function with knowledge of how our decisions affect potential outcomes. The expected utility is

$$E[\text{Utility}(\text{decision})] = \sum_{\text{possible outcomes}} p(\text{outcome} | \text{decision}) U(\text{outcome})$$

The best decision is then defined as the one that will maximize the expected utility (5).

The decision theoretic approach can be used whenever we know how decisions are related to outcomes and we know the utility of the outcomes. We can apply this framework to the example of darts playing. If we had low motor noise, we would be best off aiming at the triple 20 (Fig. 1C). In contrast, if we have a realistic value of motor noise, the best point at which to aim is not the point of maximal score but is to the lower left of the board (Fig. 1D, marked by dart). Both the behavior of advanced amateur players who have moderate motor variance and the behavior of professional players who have low motor variance are predicted by this decision theoretic approach. A range of recent studies of decision-making have analyzed situations that are analogous to playing darts (6). Such simple decisions are well predicted from the assumption that people solve their decision problems in a fashion that is close to optimal. The approach also applies to many animal behaviors: Animals need to choose to forage or rest, fight or flight, continue moving or freeze in place. Moreover, any behavior of an animal is, in some abstract way, a decision. The nervous system chooses one behavior from the set of all possible behaviors. Decision theory is thus a fundamental formalization of many problems that are solved by the nervous system and studied in neuroscience.

A number of recent studies, under the umbrella name of neuroeconomics, have started to analyze how the nervous system represents and computes with a utility function (7–9). Because

utility is central to any decision, it is important to understand how the nervous system represents reward. Utilities are important for the nervous system because it uses them to act successfully. It has been shown that the nervous system represents changes in expected utility in a way predicted by a major learning theory, reinforcement learning (10). Traditionally simple rewards, such as monetary or food incentives, are used in experiments that analyze how the nervous system represents utilities. However, insights from evolutionary biology (1) predict that many different factors will influence the utility of decisions. Indeed, it has been shown that the nervous system exhibits a reward signal when someone else who cheated in a game is punished (11). Future research will have to uncover the full complexity of how utility functions are represented and used by the nervous system.

Bayesian Statistics

Deciding seems easy: Choose the action that is associated with the highest expected utility. However, the probability of an outcome given the decision is difficult to estimate. To do so, we need to predict how the world will change until the outcome and how any decision would affect the world. This prediction can only be probabilistic because we have uncertainty about the properties of the world, stemming, for example, from noisy perception (12).

Bayesian statistics defines how uncertain pieces of information may be combined into a

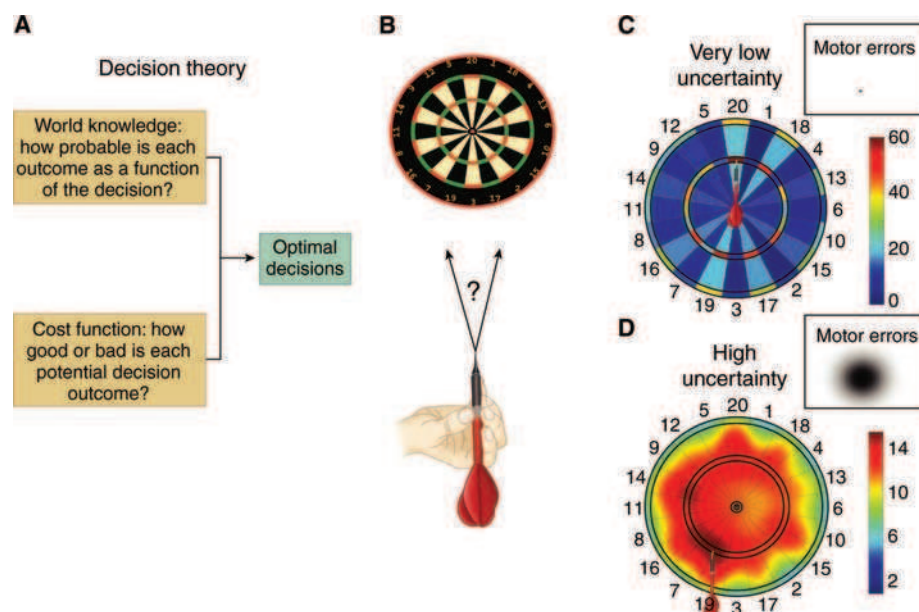


Fig. 1. Decision theory. (A) To make optimal decisions, we need to combine what we know about the world with our utility function measuring how good or bad potential outcomes are. (B) In the example of playing darts, we need to decide where to aim. (C) As a function of the aiming point, the expected score is shown for an unbelievably good darts player with almost no movement errors. (D) As in (C) but for a mediocre darts player with large motor errors.

Decision-Making

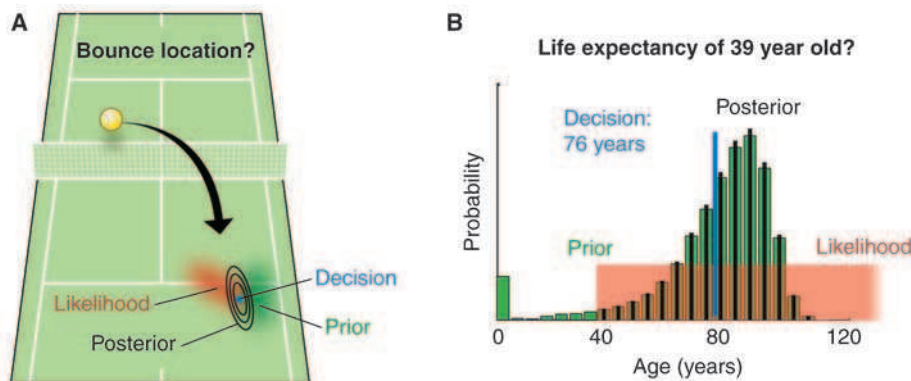


Fig. 2. (A) In the example of tennis, people need to combine what they know from before (prior, green) with what they currently see (likelihood, red). That way we can estimate the posterior (black contour lines) to make an optimal perceptual decision (blue). **(B)** Similarly if we estimate the life expectancy of a person who is 39 years old, we need to combine what we know from before (prior, histogram of lifetimes, green) with our new information (person survived 39 years, likelihood, red) to come up with an optimal estimate.

joint estimate. New information (called a likelihood) needs to be combined or integrated with information from the past (called a prior). Similar problems occur when information from several cues, for example, proprioceptive and visual, needs to be combined into a joint estimate. Bayesian decision theory (13), the use of Bayesian statistics in a decision framework, defines how our beliefs should be combined with our utility function. Because most if not all of our decisions are made in the presence of uncertainty,

understanding the way the nervous system deals with uncertainty is central to understanding its normal mode of operation.

Integration of priors and likelihoods. To calculate the probabilities of outcomes, it is often necessary to update our belief from the past (prior) with new knowledge (likelihood). For example, when we play tennis it is helpful to estimate where the ball will land. The visual system, although noisy, still provides us with an estimate or a likelihood of where the ball

will land (sketched in red in Fig. 2A). This knowledge may be combined with information obtained from experience; the positions where the ball may land are not uniformly distributed over the court. The locations may be clustered near the boundary lines, where it is most difficult to return the ball. This distribution of positions is called the prior (sketched in green in Fig. 2A). Bayes's rule states that how the probability of the ball landing at position x given our observation o (posterior) needs to be estimated as

$$\underbrace{p(x|o)}_{\text{posterior}} = \underbrace{p(x)}_{\text{prior}} \underbrace{p(o|x)}_{\text{likelihood}} / \underbrace{p(o)}_{\text{prior}}$$

Recent studies have analyzed such combinations in simple integration problems. Sensorimotor integration, force estimation, timing estimation, speed estimation, the interpretation of visual scenes, just to name a few, have been analyzed (14, 15). Together, these studies demonstrate that people intuitively combine prior knowledge with new evidence in a way predicted by Bayesian statistics.

Bayesian methods also apply to decision-making in cognitive contexts (16). What would be your guess of the life expectancy of a 39-year-old? People can use two sources of information to answer this question. They may use the prior, the distribution of lifetimes (Fig. 2B, green). They may also use the likelihood, this person must have survived the first 39 years of his or her life (Fig. 2B, red). With Bayes's rule, we can combine these two pieces of information, just as in the example of tennis (Fig. 2B, black). We thus estimate a life expectancy of about 76 years (using the negative square error as the utility function). Human participants exhibit cognitive behaviors that are close to optimal predictions. The same approach has been used to successfully predict human estimates for many other everyday cognition problems (16). People incorporate knowledge about the probability distributions into estimates in a fashion that is predicted by Bayesian decision theory.

Cue combination. Estimation will often depend on two different cues. For example, we may see and feel an object and use both senses to infer the properties of the object. Bayesian statistics allows us to solve these problems with the same mathematical framework used for the combination of prior and likelihood. A couple of recent studies have examined how subjects solve such cue combination problems. For example, the combinations of visual and auditory information and visual and tactile information, as well as within modality cue combination (e.g., texture and disparity), have been studied. In such cases, cues are combined in a fashion that is close to the optimum prescribed by Bayesian statistics (14, 15, 17, 18).

Although there is strong evidence that animals represent their degree of uncertainty and

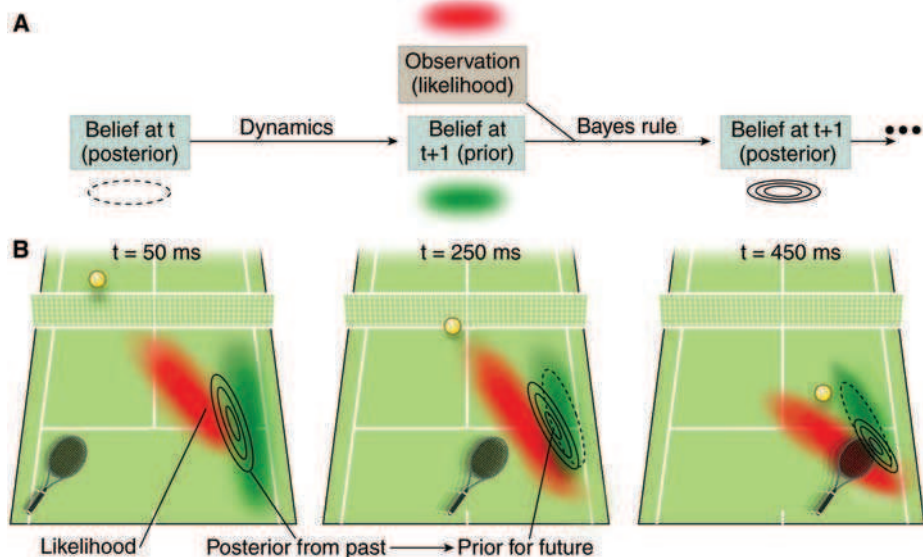


Fig. 3. Integration of information over time. (A) A diagram of a Kalman filter is shown. At any point of time t , the person has a belief about the state of the world. The person then updates this belief with a model of the dynamics of the world (e.g., gravity) to calculate the belief at the next point of time. This belief (prior) is then combined with new sensory information (likelihood) using Bayes's rule to calculate the belief at the next time step. The ellipses indicate probability distributions sketched in (B). **(B)** To estimate the position of a ball hitting the ground, people continuously update their beliefs with incoming sensory information, yielding precise estimates. The posterior of the previous time step is the prior for the new one: The dashed line indicating the previous posterior is identical to the one standard deviation line of the prior (green).

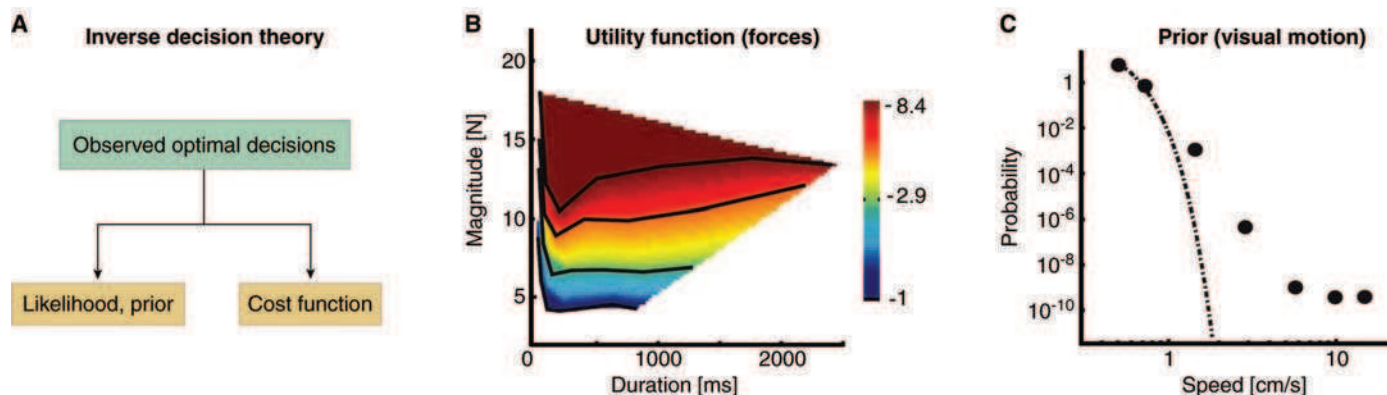


Fig. 4. Inverse decision theory. **(A)** In inverse decision theory, it is assumed that people behave optimally. From a sequence of optimal behavior, the priors, likelihoods, and utility functions people use are inferred through various compu-

tational methods. **(B)** In motor control, the utility function of producing forces of varying duration and magnitude have been calculated. **(C)** For visual decisions, people assume that small velocities are much more likely than large velocities.

combine different pieces of information into a joint estimate, there is still relatively little known about how the nervous system implements such computations. Recently, a couple of theoretical studies proposed how the nervous system may represent probabilities and combine them into estimates (19, 20). What is exciting about these studies is that they make clear testable predictions of how the nervous system may represent uncertainty. Understanding this representation is key for neuroscience because animals essentially make all their decisions in the presence of uncertainty.

Integration of information over time. The state of the world and our information about it is continually changing. In the example of tennis, we only used an estimate at a single point of time. In reality, tennis players continuously observe the ball's motion, updating their beliefs about the position where the ball will land in a continuous manner. The Kalman filter formalizes how such a process may work. The filter uses knowledge about the dynamics of the world to convert its belief about the state of the world at the previous instance of time into a belief at a future point in time. For example, in the tennis game we expect the ball to move because of its forward momentum and to accelerate because of gravity. The resulting belief (prior, Fig. 3, green) is then combined with new sensory information (likelihood, Fig. 3, red) to produce an updated belief (posterior, Fig. 3, ellipses). This way the Kalman filter updates its beliefs over time. In the case of playing tennis, such a strategy predicts that, as the ball flies, our estimate of the landing position will progressively be updated and usually become more precise. It is thus possible to make an efficient decision in situations that change continuously over time.

A range of recent studies have probed the strategies used by human participants in such situations. The way that people estimate the position of their hand in the dark is well predicted

by Kalman filtering (21). Similarly, when people balance a pole they also seem to use such strategies (22). Moreover, the way the nervous system represents changing muscle properties can be understood by assuming a Kalman filtering strategy (23, 24). These studies demonstrate that when people integrate information over time to make simple sensorimotor decisions, they seem to do so in a fashion that is consistent with the optimal Bayesian solution.

The nervous system constantly integrates information over time, and a range of new studies analyzed how it does so (25). In many such experiments, one of two stimuli is given, for example either a stimulus that moves to the right or a stimulus that moves to the left. If the stimulus is sufficiently noisy, the nervous system needs to integrate information over an extended period of time to make a good decision. Neurons were found that exhibit activities that correlate with the predicted process of optimal information integration over time. The nervous system takes into account probabilistic knowledge of potential rewards when integrating evidence for decision-making (26). The resulting models are particularly useful because they have a normative component (optimal integration of evidence) while having a straightforward descriptive component (neurons can integrate inputs over time).

Inverse Decision Theory

If people make decisions optimally, the mechanism of decision theory may be inverted: Computational techniques [e.g., (27)] are used to infer which priors, likelihoods, and utility functions the participants used to make their decisions. For example, a utility function with a few free parameters may be proposed, and the parameters may be fit to human behavior. Experimental economics (28) has extensively asked which utility functions people are optimizing. Only recently has the study of neuroscience and low-level decision-making started asking which

priors are used and which utility functions are optimized.

Many studies in neuroscience analyze motor control, an area where decision-making is key. For example, would you rather carry a 2-kg weight for 1 min or a 1-kg weight for 2 min? We intuitively make repeatable choices in such situations that are of relevance to everyday life. Such a utility function was recently inferred by using inverse decision theory (29) (Fig. 4A). This utility function is highly nonlinear in force magnitude and duration and is more complicated than previously proposed utility functions (30). This highlights a problem in decision theory: Frequently, good fits to behavior may be obtained with wrong utility functions. Inverse decision theory can thus be seen as a way of searching for violations of the assumptions made when building a decision theoretic model.

The framework of inverse decision theory also allows the analysis of which priors and which likelihoods are used by people. Studies indicate that people underestimate the speed of visual motion (31). This has been argued as the result of using a Gaussian prior for interpreting that low speeds are most likely. But why should people use a Gaussian distribution? We do not know which prior would be optimal in real life, although some recent progress in the statistics of natural scenes may lead that way (32). In a recent experiment, the prior used by human participants was measured by using inverse decision theory (33) (Fig. 4A). The prior is not Gaussian and is rather similar to an exponential function. This may inspire future experiments to characterize how the nervous system implements such a prior.

The strength of inverse decision theory in allowing for a wide range of possible utility functions is also its weakness. Inverse decision theory will always yield a utility function, likelihood, or prior for which the actually observed behavior is optimal. If the results differ from those assumed by a previous decision theoretic model, we can falsify this model. However, the

results should only be the basis of a new model if we can understand (and test) how the inferred functions derive from properties of the world. Similar problems also appear in other decision theoretic models that do not explicitly use inverse decision theory. The theoretician may fiddle with the decision theoretic model, trying different utilities, likelihoods, and priors, until there are good fits to human performance. As in all other models that explain data, overfitting is also a problem for decision theoretic models.

Inverse decision theory allows for the estimations of the used utility functions, priors, and likelihoods, which may alleviate the search for their neural representation. For example, if we know how utility functions depend on ideas of fairness when playing games, it is possible to search for brain areas that represent this aspect of the utility function (11). Searching for the neural representation of a utility function that has been proposed on theoretical grounds, but is irrelevant for human behavior, may miss important aspects of decision-making. When searching for the representation of priors or utility functions in the nervous system, it seems central to know the form of the priors and utility functions that are actually used by human participants.

Discussion

The world is complicated, and consequently so is deciding. Models in neuroscience typically analyze simple relationships between variables. However, at least our high-level behavior is characterized by structural relationships. Events in the world have causes, and we naturally interpret events in terms of cause and effect. A few studies of Bayesian statistics over the past couple of years have started to address the issue of how people may be able to infer the structure of the world (34, 35). Structure implies that not only features but their relationships play a fundamental role. This concept has long been at the heart of cognitive science. How people solve complicated real-world problems needs to be understood. A lot of recent progress in machine learning aims at inferring the structure of the world from real data (36), a process that people perform effortlessly. The study of decisions in neuroscience can draw upon advances in machine-learning to make interesting new predictions.

The decision theoretic approach may be limited in several ways. Humans should behave suboptimally for ethologically new kinds of decisions that are not repeated enough to allow for learning [but see (37)]. This may, for example, be relevant to the way people participate in lotteries (38). Using Bayes's rule to combine pieces

of information is the best mathematical solution to any information combination problem and thus always has been the best solution. Similarly, moving efficiently has always been beneficial to animals. The simple low-level properties of neural decision-making should thus be expected to be close to optimal.

Decision theory formalizes how animals should decide and thus does not directly make predictions of how the nervous system should implement the algorithm leading to such decisions (39). Countless different implementations may lead to the same optimal decision rules. However, the normative approach is not limited to decision theory. There may equally be costs and benefits to implementing algorithms in various ways. For example, having many long connections between neurons may use volume and energy and lead to slow information transmission. The idea of wiring length minimization explains well the placement of neurons in cortical maps (40) as well as within the body of the nematode *Caenorhabditis elegans* (41). Similarly, it may be argued that visual neurons should faithfully represent the world with the fewest spikes (least metabolic cost), an idea that predicts many properties of sensory representations (42). Normative ideas may even apply to cellular properties. The density of sodium channels in the squid giant axon may be understood from the idea that the squid giant axon should rapidly transmit action potentials used to allow the squid to flee from a predator (43, 44). Normative models and decision theory in particular offer ways of formalizing important problems that the nervous system needs to solve. Models in neuroscience should seek to explain the wealth of available experimental data and also incorporate knowledge of the problem solved by the system.

References

1. J. H. Barkow, L. Cosmides, J. Tooby, *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford Univ. Press, New York, 1992).
2. E. R. Weibel, C. R. Taylor, L. Bolis, *Principles of Animal Design: The Optimization and Symmorphosis Debate* (Cambridge Univ. Press, Cambridge, 1998).
3. S. J. Gould, R. Lewontin, *Proc. R. Soc. London Ser. B* **205**, 581 (1979).
4. D. Bernoulli, *Comentarii academiae scientiarum imperialis Petropolitanae* (for 1730 and 1731) **5**, 175 (1738).
5. J. v. Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
6. L. T. Maloney, J. Trommershäuser, M. S. Landy, in *Integrated Models of Cognitive Systems*, W. Gray, Ed. (Oxford Univ. Press, New York, 2006).
7. P. W. Glimcher, A. Rustichini, *Science* **306**, 447 (2004).

8. W. Schultz, *Curr. Opin. Neurobiol.* **14**, 139 (2004).
9. L. P. Sugrue, G. S. Corrado, W. T. Newsome, *Nat. Rev. Neurosci.* **6**, 363 (2005).
10. B. Seymour *et al.*, *Nature* **429**, 664 (2004).
11. D. J.-F. de Quervain *et al.*, *Science* **305**, 1254 (2004).
12. D. M. Green, J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
13. D. M. Wolpert, Z. Ghahramani, in *The Oxford Companion to the Mind*, R. L. Gregory, Ed., in press.
14. D. Knill, W. Richards, *Perception as Bayesian Inference* (Cambridge Univ. Press, New York, 1996).
15. K. P. Kording, D. M. Wolpert, *Trends Cogn. Sci.* **10**, 320 (2006).
16. T. L. Griffiths, J. B. Tenenbaum, *Psychol. Sci.* **17**, 767 (2006).
17. M. O. Ernst, H. H. Bulthoff, *Trends Cogn. Sci.* **8**, 162 (2004).
18. K. Kording *et al.*, *PLOS One* **2**, 10.1371/journal.pone.0000943 (2007).
19. W. J. Ma, J. M. Beck, P. E. Latham, A. Pouget, *Nat. Neurosci.* **9**, 1432 (2006).
20. A. Pouget, P. Dayan, R. S. Zemel, *Annu. Rev. Neurosci.* **26**, 381 (2003).
21. D. M. Wolpert, Z. Ghahramani, M. I. Jordan, *Science* **269**, 1880 (1995).
22. B. Mehta, S. Schaal, *J. Neurophysiol.* **88**, 942 (2002).
23. J. W. Krakauer, P. Mazzoni, A. Ghazizadeh, R. Ravindran, R. Shadmehr, *PLoS Biol.* **4**, e316 (2006).
24. K. P. Kording, J. B. Tenenbaum, R. Shadmehr, *Nat. Neurosci.* **10**, 779 (2007).
25. J. I. Gold, M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535 (2007).
26. T. Yang, M. N. Shadlen, *Nature* **447**, 1075 (2007).
27. A. Y. Ng, S. Russell, in *Proceedings of the Seventeenth International Conference on Machine Learning* (Morgan Kaufmann, Burlington, MA, 2000), pp. 663–670.
28. J. H. Kagel, A. E. Roth, *The Handbook of Experimental Economics* (Princeton Univ. Press, Princeton, NJ, 1995).
29. K. P. Kording, I. Fukunaga, D. Wolpert, *PLoS Biol.* **2**, e330 (2004).
30. E. Todorov, *Nat. Neurosci.* **7**, 907 (2004).
31. Y. Weiss, E. P. Simoncelli, E. H. Adelson, *Nat. Neurosci.* **5**, 598 (2002).
32. B. Y. Betsch, W. Einhauser, K. P. Kording, P. Konig, *Biol. Cybern.* **90**, 41 (2004).
33. A. A. Stocker, E. P. Simoncelli, *Nat. Neurosci.* **9**, 578 (2006).
34. J. B. Tenenbaum, T. L. Griffiths, *Behav. Brain Sci.* **24**, 629 (2001).
35. J. B. Tenenbaum, T. L. Griffiths, C. Kemp, *Trends Cogn. Sci.* **10**, 309 (2006).
36. A. Yuille, D. Kersten, *Trends Cogn. Sci.* **10**, 301 (2006).
37. J. J. Koehler, *Behav. Brain Sci.* **19**, 1 (1996).
38. D. Kahneman, A. Tversky, *Econometrica* **47**, 313 (1979).
39. D. Marr, *Vision: A Computational Approach* (Freeman, San Francisco, 1982).
40. D. B. Chklovskii, A. A. Koulakov, *Annu. Rev. Neurosci.* **27**, 369 (2004).
41. B. L. Chen, D. H. Hall, D. B. Chklovskii, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4723 (2006).
42. B. A. Olshausen, D. J. Field, *Curr. Opin. Neurobiol.* **14**, 481 (2004).
43. A. L. Hodgkin, *Philos. Trans. R. Soc. London Ser. B* **270**, 297 (1975).
44. T. D. Sangrey, W. O. Friesen, W. B. Levy, *J. Neurophysiol.* **91**, 2541 (2004).

10.1126/science.1142998

4. S. Ito, V. Stuphorn, J. W. Brown, J. D. Schall, *Science* **302**, 120 (2003).
5. K. Shima, J. Tanji, *Science* **282**, 1335 (1998).
6. J. O'Doherty, M. L. Kringelbach, E. T. Rolls, J. Hornak, C. Andrews, *Nature Neurosci.* **4**, 95 (2001).
7. M. Ullsperger, D. Y. von Cramon, *J. Neurosci.* **23**, 4308 (2003).
8. C. B. Holroyd, M. G. H. Coles, *Psychol. Rev.* **109**, 679 (2002).
9. C. B. Holroyd, J. T. Larsen, J. D. Cohen, *Psychophysiology* **41**, 245 (2004).
10. H. Gemba, K. Sasaki, V. B. Brooks, *Neurosci. Lett.* **70**, 223 (1986).
11. M. Ullsperger, D. Y. Von Cramon, *Cortex*, in press.
12. N. Picard, P. L. Strick, *Cereb. Cortex* **6**, 342 (1996).
13. C. B. Holroyd, S. Nieuwenhuis, R. B. Mars, M. G. H. Coles, in *Cognitive Neuroscience of Attention*, M. I. Posner, Ed. (Guilford, New York, in press).
14. W. J. Gehring, B. Goss, M. G. H. Coles, D. E. Meyer, E. Donchin, *Psychol. Sci.* **4**, 385 (1993).
15. M. Falkenstein, J. Hoormann, S. Christ, J. Hohnsbein, *Biol. Psychol.* **51**, 87 (2000).
16. W. Schultz, *Neuron* **36**, 241 (2002).
17. C. B. Holroyd et al., *Nature Neurosci.* **7**, 497 (2004).
18. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, *Psychol. Rev.* **108**, 624 (2001).
19. N. Yeung, M. M. Botvinick, J. D. Cohen, *Psychol. Rev.*, in press.
20. Materials and methods are available as supporting material on Science Online.
21. R. Hester, C. Fassbender, H. Garavan, *Cereb. Cortex* **14**, 986 (2004).
22. The majority of activations fall into the border zone between areas 8, 6, and 32, with some extension into area 24. Recent research in nonhuman primates seems to suggest a functional-anatomical dissociation of regions subserving pre-response conflict monitoring from structures sensitive to errors and omission of reward (1, 4). Although in humans this view is still under debate (11, 13, 21), the present meta-analysis does not provide unequivocal evidence for or against such a dissociation. Activations related to pre-response conflict and uncertainty occur more often in area 8 and less often in area 24 than do signal increases associated with errors and negative feedback (area 8, 32.5% versus 9.7%; area 24, 7.5% versus 25.8%), supporting the dissociation view. However, both groups of activations cluster primarily in area 32 (pre-response, 42.5%; error, 41.9%), suggesting that pre- as well as post-response monitoring processes share at least one underlying structure. It seems that the currently available spatial resolution in fMRI, in conjunction with anatomical variability and differences in scanning and preprocessing methods between studies, limit the ability to resolve this debate about a possible dissociation in the range of 10 mm or less.
23. T. Paus, *Nature Rev. Neurosci.* **2**, 417 (2001).
24. H. D. Critchley et al., *Brain* **126**, 2139 (2003).
25. E. K. Miller, J. D. Cohen, *Annu. Rev. Neurosci.* **24**, 167 (2001).
26. A. R. Aron, T. W. Robbins, R. A. Poldrack, *Trends Cogn. Sci.* **8**, 170 (2004).
27. D. Badre, A. D. Wagner, *Neuron* **41**, 473 (2004).
28. S. A. Bunge, K. N. Ochsner, J. E. Desmond, G. H. Glover, J. D. E. Gabrieli, *Brain* **124**, 2074 (2001).
29. M. Brass, D. Y. von Cramon, *J. Cogn. Neurosci.* **16**, 609 (2004).
30. K. R. Ridderinkhof et al., *Science* **298**, 2209 (2002).
31. K. R. Ridderinkhof, S. Nieuwenhuis, T. R. Bashore, *Neurosci. Lett.* **348**, 1 (2003).
32. J. G. Kerns et al., *Science* **303**, 1023 (2004).
33. H. Garavan, T. J. Ross, K. Murphy, R. A. Roche, E. A. Stein, *Neuroimage* **17**, 1820 (2002).
34. G. Bush et al., *Proc. Natl. Acad. Sci. U.S.A.* **99**, 523 (2002).
35. J. O'Doherty, H. Critchley, R. Deichmann, R. J. Dolan, *J. Neurosci.* **23**, 7931 (2003).
36. R. Cools, L. Clark, A. M. Owen, T. W. Robbins, *J. Neurosci.* **22**, 4563 (2002).
37. J. F. Bates, P. S. Goldman-Rakic, *J. Comp. Neurol.* **336**, 211 (1993).
38. M. Petrides, D. N. Pandya, *Eur. J. Neurosci.* **11**, 1011 (1999).
39. L. Koski, T. Paus, *Exp. Brain Res.* **133**, 55 (2000).
40. K. Fiehler, M. Ullsperger, D. Y. von Cramon, *Eur. J. Neurosci.* **19**, 3081 (2004).
41. W. J. Gehring, R. T. Knight, *Nature Neurosci.* **3**, 516 (2000).
42. M. I. Posner, G. J. DiGirolamo, in *The Attentive Brain*, R. Parasuraman, Ed. (MIT Press, Cambridge, MA, 1998), pp. 401-423.
43. K. Matsumoto, K. Tanaka, *Science* **303**, 969 (2004).
44. K. Matsumoto, K. Tanaka, *Curr. Opin. Neurobiol.* **14**, 178 (2004).
45. This research was supported by a TALENT grant (E.A.C.) and a VENI grant (S.N.) of the Netherlands Organization for Scientific Research and by the Priority Program Executive Functions of the German Research Foundation (M.U.). Helpful comments by S. Bunge are gratefully acknowledged.

Supporting Online Material
www.sciencemag.org/cgi/content/full/306/5695/443/DC1
 Materials and Methods
 Table S1
 References

REVIEW

Neuroeconomics: The Consilience of Brain and Decision

Paul W. Glimcher^{1*} and Aldo Rustichini²

Economics, psychology, and neuroscience are converging today into a single, unified discipline with the ultimate aim of providing a single, general theory of human behavior. This is the emerging field of neuroeconomics in which consilience, the accordance of two or more inductions drawn from different groups of phenomena, seems to be operating. Economists and psychologists are providing rich conceptual tools for understanding and modeling behavior, while neurobiologists provide tools for the study of mechanism. The goal of this discipline is thus to understand the processes that connect sensation and action by revealing the neurobiological mechanisms by which decisions are made. This review describes recent developments in neuroeconomics from both behavioral and biological perspectives.

The full understanding of utility will come from biology and psychology by reduction to the elements of human behavior followed by a bottom-up synthesis, not from the social sciences by top-down inference and guesswork based on intuitive knowledge. It is in biology and psychology that economists and social scientists will find the

premises needed to fashion more predictive models, just as it was in physics and chemistry that researchers found the premises that upgraded biology. (p. 206) (1)

Consider the famous St. Petersburg paradox (2). Which of the following would you prefer, \$40 or a lottery ticket that pays according to the outcomes of one or more fair coin tosses: heads you get \$2 and the game ends, tails you get another toss and the game repeats, but now if the second toss lands heads up you get \$4, and so on. If the *n*th toss is the first to land heads up, you get

2^{*n*} dollars. The game continues, however long it takes, until the coin lands heads up. We can assess the average objective, or expected, value of this lottery by multiplying the probability of a win on each flip by the amount of that win:

$$\begin{aligned}\text{Expected value} &= (0.5 \times 2) + (0.25 \times 4) + \\ &\quad (0.125 \times 8) \dots \\ &= 1 + 1 + 1 + \dots\end{aligned}$$

This simple calculation reveals that the expected value of the lottery is infinite even though the average person is willing to pay less than \$40 to play it. How could this be?

For an economist, any useful explanation must begin with a set of assumptions that renders behavior formally tractable to coherent theoretical and mathematical analysis. Economists therefore explain this behavior by assuming that the desirability of money does not increase linearly, but rather grows more and more slowly as the total amount at stake increases. For example, the desirability of a given amount might be a power function

¹Center for Neural Science, New York University, New York, NY 10003, USA. ²Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA.

*To whom correspondence should be addressed. E-mail: glimcher@cns.nyu.edu

of that amount, as shown by the black line in Fig. 1. A decision-maker for whom the subjective value, or utility, of money grew in this fashion would then determine the desirability, or expected utility, of the St. Petersburg lottery by multiplying the probability of a win on each flip by the utility of the amount won on that flip, and thus he might well be willing to pay less than \$40 to play this game.

From the point of view of a psychologist attempting to understand and explain this same phenomenon, it is the nature of risk aversion and the psychological mechanisms that this set of preferences reveals that become the subject of explanatory study. The psychological mechanism that accounts for risk aversion in human subjects, for example, has been shown to be more sensitive to monetary losses than to monetary gains. Further, psychologists have suggested that subjective utilities are computed with regard to somewhat arbitrary and idiosyncratic monetary reference points, or frames, set by yet other psychological processes (3). Psychologists use observations like these to argue that human choosers are endowed with a particularly strong fear of losses and that they weigh the merits of all possible gains and losses relative to a psychological benchmark: The psychological approach seeks empirically to describe minimally complex behavioral tendencies, modules, or heuristics that can account for the actions of human choosers.

A traditional neurobiological perspective uses yet another approach: A hungry bird is shown a tray that contains five millet seeds and repeatedly permitted to fly to the tray and eat the seeds. At a neurobiological level, the study of this behavior begins with the assumption that the visual stimulus of the five seeds must somehow propagate through the sensory system of the animal to trigger activation in orienting circuits that move the bird to the seeds. Next, the same bird is permitted to fly to a second tray covered by a piece of paper. When the bird displaces the cover, half of the time it reveals 12 seeds and half of the time it reveals nothing. Mechanistically, the visual stimulus must again trigger an orienting response, and presumably in this case the strength with which visual signals connect synaptically to the orienting circuits reflects both the number of seeds that the bird earns and the likelihood that seeds will be found under the paper. Lastly, both trays are presented, and the bird is observed to fly toward the tray that may contain 12 millet seeds. A standard neurobiological explanation (4, 5) presumes that under these circumstances the two different behavioral circuits compete. In this case the synapses that elicit an orienting response to the covered tray are stronger and thus control

behavior. The neurobiological explanation specifies the minimal neural circuitry required to account for the observed behavior of the bird.

What is striking about explanations of choice behavior by economists, psychologists, and neurobiologists is the different levels at which they operate. The economic approach attempts to describe globally all choice behavior with a single logically consistent formalism. The psychological approach examines the ways in which subjective and objective estimates of value differ and posits psychological modules that can account for these observed behavioral preferences. The neurobiological explanation starts with the simplest possible neural

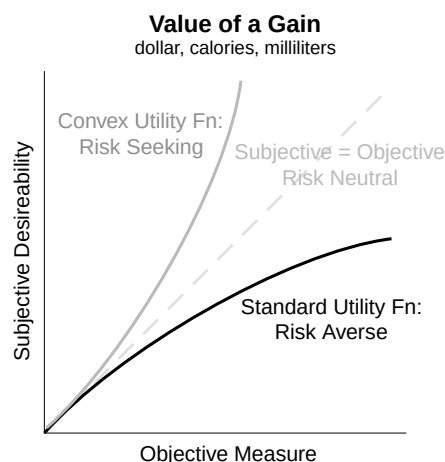


Fig. 1. Bernoulli's notion of subjective value or utility. The black line plots the typical relationship between objective and subjective valuations of an action. As the objective value of a gain increases, the subjective desirability, or utility, grows more slowly. Bernoulli demonstrated that this relationship could account for the observation that humans are typically risk-averse. The solid gray line plots a condition in which subjective value grows more quickly than objective value, a preference structure that would yield risk-seeking behavior.

circuitry that can account for the simplest measurable elements of behavior. It seems obvious that these different levels of explanation should be linked, but how can such a linkage be accomplished? We argue that a unified explanation of decision-making is not only possible but has recently begun and that, when the linkage between these three levels of explanation has matured, a new, more powerful decision science rooted in a neuroeconomic approach will have been developed.

A second claim we will make is that once this reconstruction of decision science is completed, many of the most puzzling aspects of human behavior, aspects that economic theory, psychological analysis, or neurobiological deconstruction have failed to

explain, will become formally and mechanistically explicable. The claim is, in essence, that a decision science that simultaneously engaged all three approaches would be more heavily constrained and at the same time would have much greater explanatory power than do any of these three approaches operating alone. We will see examples of how this synthetic approach would operate in principle and early attempts at synthetic solutions below.

This reconstruction of the study of decision is also going to be the appropriate basis for a more ambitious theory that explains not just how we make decisions but why. That such an explanation is necessary and possible is indicated by the fact that fundamental features of decision making are common to many species. For example, risk aversion as shown by the St. Petersburg paradox has been described in many species. Studies of birds making choices in risky environments produce a behavior best described by a utility function (Fig. 1) (6, 7). We know that humans and birds deviated from a common reptilian ancestor at least 200 million years ago, but this basic function for choice has remained essentially unchanged. Such commonalities make a clear suggestion: A utility function of this type probably is an efficient and evolved feature of vertebrate choice. For example, Robson (8) provides a justification of why a utility function might be an evolutionary optimal response to changing environments. Just as information theory was used by Barlow (9) to explain why animals as diverse as horseshoe crabs and cats use similar encoding schemes in their visual systems, an economic theory that relates utility to Darwinian fitness must serve as an overarching tool for understanding vertebrate choice behavior.

Linking the Decision Sciences

Subjective desirability. The central concept in modern economic theory is the notion of subjective utility: Preferences must be described as subjective properties of the chooser. Surprisingly, the notion that preferences are represented in the nervous system, that these preferences are subjective, and that they guide the production of action has only recently entered the neurobiological mainstream. We believe that this has been a critical flaw in neurobiological studies, because it is essential that economics, psychology, and neuroscience acknowledge a common phenomenological base to achieve a reductive unification of the decision sciences. The concepts that guide the behavioral study of decision-making must also guide the mechanistic study of that process.

In part, this preference-free approach may have arisen from neurobiology's roots in the stimulus-response physiology of the

early twentieth century (10). Working from Descartes' philosophy, Sherrington (11) proposed that physiologists should work to link stimulus and response directly through what Pavlov (12) would later call a "definite nervous path." Scientists like Sherrington and Pavlov proposed that it was the role of neuroscience to chart these stimulus-response connections through the nervous system.

A critical step beyond this initial framework was a recent effort to explain more complicated behaviors and to focus on actions for which deterministic sensory-to-movement mapping approaches were insufficient. Newsome and his colleagues (13, 14) made that step in the late 1980s when they examined perceptual decision-making by monkeys viewing ambiguous sensory stimuli. In those experiments, monkeys stared at a display of chaotically moving spots of light. On training trials, a subset of the spots moved coherently in a single direction, whereas the remaining spots moved randomly (15). The direction of this coherent motion indicated which of two possible saccadic eye movements would yield a fruit juice reward, and at the end of each trial animals were free to make a saccade. If they made the correct movement, they then received the reward. On a critical subset of trials, however, monkeys viewed displays in which none of the dots moved in a single coherent direction, and thus the display provided no information from which the location of the rewarded eye movement could be deduced. Under these conditions, Newsome and his colleagues found that the firing rates of single neurons in the middle temporal visual area (area MT) were still correlated with the behavior of the animals, even when that behavior could not be predicted from the properties of the stimulus. Newsome and his colleague Shadlen's subsequent studies revealed the basic neurobiological substrate for perceptual decision-making and showed convincingly that this circuit could not be modeled simply as a single "definite nervous path" from stimulus to response (16).

This work, in turn, accelerated studies of the posterior parietal cortex, an area interposed between many of the sensory circuits and motor circuits of the primate brain, which appeared to play a critical role in the perceptual decision-making Newsome has studied (17, 18). Platt and Glimcher (19) made an important advance when they extended Newsome's approach by proposing that posterior parietal cortex might play a role in decision-making in an economic

sense and that it might encode the desirabilities of making particular movements.

In Platt and Glimcher's experiments, trained rhesus monkeys were allowed to participate in repeated rounds of a simple lottery while the activity of nerve cells in the posterior parietal cortex was monitored. At the beginning of each round, two yellow spots were illuminated on a screen, one to the left and one to the right of where the monkey was looking. This began the lottery phase of the round, a period during which the monkey did not know whether the left or right light would be offered as a prize at the end of that round. At the end of this phase, a third light changed color to red or green, indicating which of the two initial lights had been randomly selected to yield a fruit juice reward on that particular round. The monkey received the fruit juice if he oriented to the selected light at the end of the round. While monkeys played hundreds of rounds of this game, Platt and Glimcher systematically varied either the relative probabilities that the left or right lights would be selected at

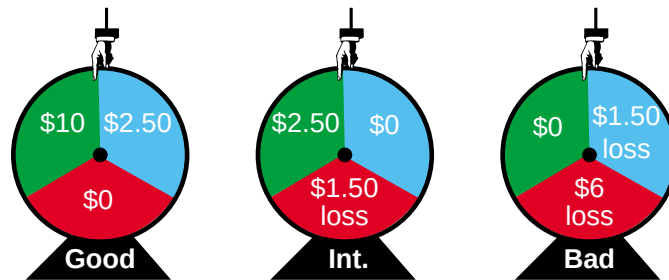


Fig. 2. The three lotteries used in the Breiter and colleagues experiment.

the end of each round or the size of the reward associated with each. These two variables were selected because economic theories assess desirability by combining the value and likelihood of gain in some subjective manner. Platt and Glimcher found that some parietal neurons did indeed encode the value and likelihood of reinforcement during the lottery phase of each round. Under these conditions, the brains of the monkeys explicitly encoded something very much like the economically defined expected value or expected utility of each light in this simple lottery task.

Subsequent studies of human decision-making using functional magnetic resonance imaging (fMRI) have yielded similar conclusions. Knutson and colleagues (20) have shown, for example, that activity in the human striatum is correlated with the magnitude of the monetary reward subjects earn during lotteries, and Paulus and colleagues (21) have shown a similar result in the human posterior parietal cortex. In a particularly interesting study, Breiter and col-

leagues (22) (Fig. 2) presented human subjects on sequential rounds with one of three possible lotteries. In lottery one (the good lottery), they faced equal chances of winning \$10, \$2.50, or \$0. In lottery two (the intermediate lottery), they faced equal chances of winning \$2.50, winning \$0, or losing \$1.50. In lottery three (the bad lottery), they faced equal chances of winning \$0, losing \$1.50, or losing \$6. At the beginning of each round, the subjects were told which lottery they would be playing, and the average activity in many brain areas was simultaneously measured. After that measurement was complete, the lottery was actually played and the humans were then told how much real money they had earned on that round. Importantly, all three of these lotteries present a one-third possibility of winning \$0, but they do so under different conditions. In the good lottery, winning \$0 is the worst possible outcome, whereas in the bad lottery it is the best. The psychologists Kahneman and Tversky (23) have shown that, when a human participates in this good lottery, they find winning \$0 to be an intensely negative outcome whereas when a human participates in the bad lottery, they find winning \$0 to be a positive outcome; subjective utilities are computed with regard to a reference frame. Breiter and colleagues found that the activity of the sublenticular extended amygdala encoded the desirability of each lottery, taking into account this behaviorally described framing effect.

Other recent neurobiological studies have revealed yet other neurally encoded variables that include the log likelihood that a given eye movement will result in a reward (24, 25), the very closely related integral of perceptual signals indicating which saccade will be rewarded in the Newsome task (26, 27), the average rate at which a saccade has been rewarded in the recent past (28), the instantaneous likelihood, or hazard, that a reinforced saccade will be instructed (29), and combinations of these variables (30).

Strategic thinking. All of these results suggest that classical utility theory can be used as a central concept for the study of choice in economics, psychology, and neuroscience. In the middle of the twentieth century, however, economists pushed utility theory beyond this boundary, enhancing it to include the study of the strategic interactions which arise when decision makers confront intelligent opponents. Extending the concept of subjective utility, VonNeumann and Morgenstern (31) and Nash (32, 33) developed a formal utility-based economic approach in the theory of

games. Recently, Lee and his colleagues (34) and Dorris and Glimcher (35) have begun to link the neurobiological corpus to this literature by examining the activity of single neurons in awake-behaving monkeys engaged in decision-making during strategic conflicts. In Dorris and Glimcher's study, two opponents face each other, an employer and an employee. On each round of the game the employee must decide whether to go to work, in which case he earns a fixed wage, or whether to shirk, in hopes of earning his wage plus a bonus. The goal of the employee is simply to maximize his gains in terms of salary and bonus. The employer, on the other hand, must decide between trusting his employee to arrive for work or spending money to hire an inspector who can actually check and see whether the employee arrived for work that day. The goal of the employer is to spend as little as possible on inspections while maximizing the employee's incentive to work.

The inspection game is of particular interest to game theorists and economists because rational strategies for utility maximization during this strategic conflict lead to predictable outcomes according to an equilibrium theory originally developed by John Nash in the 1950s. Nash (36) equilibrium theory describes how, when the cost of inspection to the employer is set high, the efficient strategy for both players converges on a solution in which the employee manages to shirk fairly often. Conversely, a low inspection cost to the employer defines a theoretical equilibrium solution in which shirk rates are low. One of the fundamental insights this formal analysis reveals is that at a mixed strategy equilibrium, a situation in which a rational player should distribute his actions amongst two or more alternatives in an unpredictable fashion, the desirability of the two or more actions in equilibrium must be equivalent. The Nash approach argues, essentially, that a behavioral equilibrium occurs when the desirability of working and shirking are rendered equal by the behavior of one's opponent irrespective of how often that equilibrium requires that one work. When Dorris and Glimcher examined the activity of neurons in the posterior parietal cortex of monkeys playing the inspection game, they found neurons that carried a signal that behaved like relative expected utility. When the monkey's behavior was well predicted by the Nash equations, neural activity was fixed at a single level irrespective of the frequency with which the monkey

chose to make a particular response, even though these same neurons were strongly modulated by changes in the value of actions during lottery tasks.

Research on human-human strategic interactions that are well described by classical game theory are also now under way in a number of laboratories (37). Like the earlier fMRI studies of simple decision-making tasks, these experiments are also beginning to shape the common ground between economics, psychology, and neuroscience. Taken together, these findings suggest that at least under some circumstances decisions may actually be made in the primate neuro-architecture in a manner long suspected by economists and now being actively analyzed by psychologists and neuroscientists: Neural circuits may compute and represent the desirability of making a response. Economics, psychology, and neuroscience do seem to be

sequent puzzles and paradoxes, however, have plagued this solution. In Ellsberg's (1961) paradox (38) (Fig. 3) you are presented with an urn, and you are told that it contains 90 balls. Of these, 30 are blue, and 60 are either red or yellow; any proportion is possible. You are then offered a choice between a lottery that pays \$100 if a blue ball is drawn (a 1/3 probability) and one that pays \$100 if a red ball is drawn. The probability of a red draw is unspecified or ambiguous: It is a choice between an event with a known probability and an event with an unknown probability. Under these circumstances, people typically choose the first lottery, which wins if a blue ball is drawn. According to expected utility theory they could only do so if they believe that there are fewer than 30 red balls in the urn or, equivalently, that there are more than 30 yellow balls. Then (before any balls are actually drawn, but with the same urn standing in front of you) you

are asked to choose again, this time between a lottery that pays \$100 on either blue or yellow and one that pays \$100 on either red or yellow. Now the likelihood of winning is clear in the second case (a 2/3 probability of winning \$100) but unclear in the first case (a probability between 1/3 and 1). People this time typically choose the second lottery. The first lottery seems less attractive, because there might be too few yellow balls. Is there anything wrong with this behavior? If expected utility

theory is correct, then there certainly is: You cannot think that there are too few and too many yellow balls in the urn at the same time.

The Ellsberg paradox is just one of many demonstrations presented in the last half of the twentieth century that were considered formal falsifications of expected utility theory. An even earlier example is Allais' paradox (39), based on the idea that a certain outcome may be perceived as more desirable, in a qualitatively different way, than any random outcome, even if very likely (40). These examples proved that expected utility theory as originally proposed could not be globally correct; at best it could only predict choices under some circumstances. This has led economists and social psychologists both to attempt modifications to expected utility theory and to replace it outright. Both the modifications and replacements have provided important and economically powerful insights into choice behavior but have not yet provided a global theory of

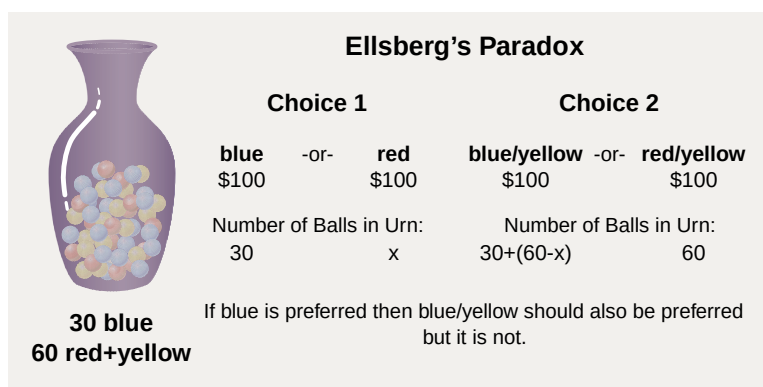


Fig. 3. Ellsberg's paradox. If blue is preferred in choice one, then blue/yellow should logically be preferred in choice two. Surprisingly, this is rarely the case.

converging around a common conceptual framework. All three disciplines are beginning to acknowledge that decision-making involves the representation of subjective desirabilities. The challenge that this convergence around a single concept poses, however, is to leverage the intersection of these three disciplines to explain choice behavior that cannot be described with the common framework of utility theory; these are classes of behaviors which have stymied traditional economics and which have lain far beyond the reach of traditional neuroscience. If it is to be of value, the goal of a unified decision science will have to be to use all three sets of approaches simultaneously to gain traction in this new territory.

Beyond Classical Concepts

Choice under risk. As we have seen, the introduction of the concept of expected utility solved the St. Petersburg's puzzle and formed the core of neoclassical economics. Sub-

choice that can truly replace expected utility theory.

The emerging discipline of neuroeconomics offers a new strategy both for testing existing models of all types and for developing new models with empirical techniques. If we succeed in understanding mechanistically how choices that violate expected utility theory are made at a neural level, then a new global theory of choice will be developed. To that end, a number of laboratories are now beginning to reexamine the conditions under which expected utility theory fails.

The reason that expected utility theory fails under some conditions may be that choosers use more than one evaluative mechanism at a neurobiological level (41). For example, in Dickhaut *et al.* (42), the processes involved when a certain outcome is one of the options are different from those involved when only random outcomes are at stake, providing an explanation of the Allais' paradox cited above.

Under many conditions these mechanisms may work together to yield choices similar to those predicted by expected utility theory but may produce odd results when used in isolation, in novel combinations, or in situations for which they are ill suited. Recent work by Damasio and colleagues [for example, (43)] on the class of behavioral paradoxes from which the Ellsberg example is drawn seem to support this conclusion. These studies suggest that an ambiguity-sensitive mechanism associated with the expression of emotion may reside, at least in part, in the ventromedial prefrontal cortex (VMPFC) (Fig. 4) and may be responsible for choice under some but not all conditions. These researchers and others have shown that patients with damage to this area have an impaired ability to make some classes of decisions and have difficulties planning their work and choosing friends. Further, the actions these individuals do elect to pursue often lead to financial as well as personal losses. Yet despite these specific failures, patients with damage to the VMPFC show normal performance on multiple-choice tests of intelligence.

These observations and others like them have led Damasio to propose that the inability of patients with VMPFC lesions to make advantageous decisions under some circumstances is caused by damage to an emotional mechanism that stores and signals the value of future consequences of an action, the somatic marker hypothesis. The hypothesis proposes that, because they lack this emotional mechanism, the patients

must rely on other brain mechanisms that achieve a different analysis of the numerous and often conflicting options involving both immediate and future consequences. This other mechanism, operating alone, is hypothesized to produce decisions that are less efficient and slower than those produced by a normal, intact, system.

The importance of the emotion-related VMPFC for regular decision-making has been confirmed by experiments where subjects were asked to make choices among a group of alternatives that carry a monetary reward (typically by selecting one card at a time from four different decks of cards), but for which the probability of reward is unspecified (44). This is precisely the ambiguous situation that produces Ellsberg's paradox. Under these conditions, patients with VMPFC lesions seem to lack

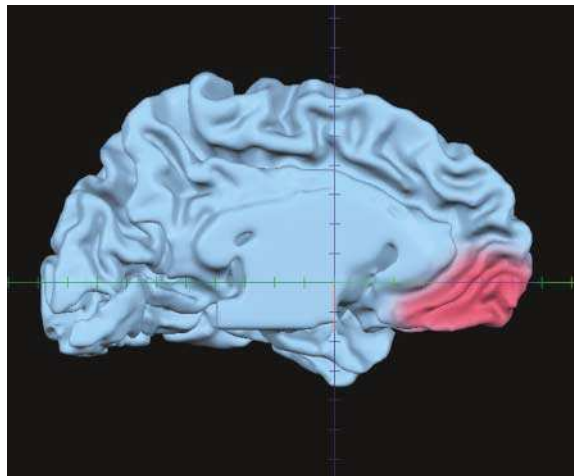


Fig. 4. Medial view of the left half of a human brain, with the front of the brain on the right side of the image. The human ventromedial prefrontal cortex is shown in red.

an aversion to ambiguity or losses that normal subjects have, an aversion that may be quite advantageous under many conditions. Further support for this hypothesis comes from brain imaging studies. For example, O'Doherty *et al.* (45) have shown that the VMPFC is relatively more active when human subjects are actively learning about the availability of rewards and punishments during one of these ambiguous choice tasks.

The process may be very different, however, when subjects simply choose between options without any feedback or learning taking place at the same time. For example, Rustichini *et al.* (46) asked normal subjects to make choices among ambiguous lotteries, risky lotteries, and certain outcomes while their brain activity was monitored. Subjects were paid for the outcome of their choices, but the outcome was communicated only after the experiment was over.

Under these conditions, the VMPFC did not show any activation; it was actually less active when choices were being made than when subjects waited between trials. These results suggest that emotional circuits may be important in learning and processing information, rather than in selecting among alternatives.

Together, these data may begin to explain, in a mechanistic way, how information is analyzed when at least one class of behavior which is not predicted by the expected utility theory is produced. The process of learning and evaluating feedback may involve emotion-related areas. Ambiguity aversion, whether advantageous or disadvantageous in a particular situation, may become explicable as we learn more about the computations that brain areas like the VMPFC perform.

Strategic cooperation. As with the Ellsberg paradox, challenges have also been raised recently to classical game theory. In a path-breaking study, Guth *et al.* (47) analyzed the behavior of subjects playing the ultimatum game. In this game, a first player, the proposer, has \$10 to split with a second player. He can offer any amount between zero and \$10. The second player is informed of the offer and can accept or refuse. If she accepts, the split is made. If she refuses, both players get nothing. The prediction of a restrictive concept of game theory, the subgame perfect equilibrium, is that for any positive amount offered by the proposer, the second player knows that she faces a choice between gaining nothing (if she refuses the offer) or something (if she accepts). The proposer should therefore always offer the minimum possible split to player two, who should always accept. Contrary to this prediction, the robust experimental finding is that low offers (typically \$2 or even \$3) are consistently refused. The second player appears to prefer, under these conditions, to gain nothing. Anticipating this, proposers typically avoid low offers.

Although expected utility theorists have proposed some explanations for this behavior, it may well be that by analyzing the neural circuits active during the ultimatum game we may be able to both explain the causes of this behavior and to predict it. Studying the ultimatum game in subjects undergoing brain scans, Sanfey *et al.* (48) found that offers refused by the second players activated specific brain circuits in those players, and interestingly these brain circuits are also associated with emotional arousal: the anterior insula (AI, associated with disgust, both physical and emotional), the dorsolateral prefrontal

cortex (DLPFC, associated with goal maintenance and executive control), and the anterior cingulate cortex (ACC, associated with detection of cognitive conflict). Also significant is the correlation of activation with choices: An activation of the AI is positively correlated with rejection, suggesting that an emotional arousal associated with a low offer is correlated with rejection. The overall picture is that offers we might consider unfair may activate emotional circuits of the brain involved in the decision to reject an offer. If we can come to more fully understand how these circuits reach this conclusion, then a behavior that was difficult for classical game theory to predict may become fully explicable with the synthetic approach that neuroeconomics provides.

A similar line of investigation has examined interplayer cooperation during single rounds of the trust game. In this game, two players move sequentially. The first player can decide to transfer a sum of money out of an initial endowment that she receives into an investment pool that immediately triples in value. The second player then gains control of the investment and can divide it between the two players in any way he chooses. In this game, the only Nash equilibrium choice for the first player is to transfer nothing into the investment. Were she to make any transfer, the second player should take all of the money for himself. However, in real experiments the first player typically does transfer a significant amount into the investment, and the second player reciprocates by returning part of the pool. McCabe *et al.* (49) had subjects play the trust game both against a human opponent and against a computer program which, they were told, would play a human-like strategy. Under these conditions McCabe and colleagues found that subjects were more likely to cooperate with real humans than with computers and that cooperators have a significantly different brain activation in the two conditions. Cooperation is associated with activation of the anterior paracingulate cortex, a brain region associated with (50, 51) interpreting and monitoring the mental state of others.

Although these studies are in their early stages, they suggest the existence of specific brain components that make specialized contributions to decision making. The challenge that these studies face is to derive detailed computational models of the neural mechanisms, which will make neuroeconomic models broadly predictive as well as explanatory.

Summary

Economics, psychology, and neuroscience are converging into a single, unified field

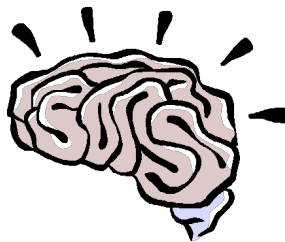
aimed at providing a theory of human behavior. In this enterprise, the method and the standard set by neuroscience is the final goal: a reconstruction of the process and mechanism that goes from a stimulus presented to the subject to his final action in response. Economics provides the conceptual structure and the object of the analysis. In this emerging view, people are seen as deciding among options on the basis of the relative desirability of each option. This is true when they are in isolation as well as when they are in strategic (interaction with few persons) and market (interaction with a large number) environments. The recent research we have been surveying describes how desirability is realized as a concrete object, a neural signal in the human and animal brain, rather than as a purely theoretical construction. Desirability is computed and is represented in the brain, and we now have the means to test, measure, and represent this activation.

But the complete reconstruction of the decision process, and hence of human behavior, is not going to be easy, because two of the cornerstones of economic analysis, subjective utility theory and Nash equilibrium, provide, even from the descriptive point of view, an incomplete picture. For example, desirability as represented by the simple economic formalism of expected utility may be appropriate only in simple conditions, where ambiguity is excluded. A more general notion is needed and, as we have seen, is beginning to be investigated and developed by psychologists and economists working together. The goal of the emerging neuroeconomic program will have to be a mechanistic, behavioral, and mathematical explanation of choice that transcends the explanations available to neuroscientists, psychologists, and economists working alone. Although it is unclear today how complete this explanation will ultimately be, neuroeconomic approaches have already begun to yield substantial fruit and to fuse natural and social scientific approaches to the study of human behavior.

References and Notes

1. E. O. Wilson, *Consilience* (Knopf, New York, 1998).
2. D. Bernoulli, *Econometrica* **22**, 23 (1954).
3. D. Kahneman, P. Slovic, A. Tversky, Eds, *Judgement Under Uncertainty: Heuristics and Biases* (Cambridge Univ. Press, Cambridge, 1982).
4. C. S. Sherrington, *The Integrative Action of the Nervous System* (Scribner, New York, 1906).
5. I. P. Pavlov, *Conditioned Reflexes an Investigation of the Physiological Activity of the Cerebral Cortex* (Oxford Univ. Press, Oxford, 1927).
6. T. Caraco, S. Martindale, T. Whittam, *Anim. Behav.* **28**, 820 (1980).
7. A. Kacelnik, M. Bateson, *Trends Cogn. Sci.* **1**, 304 (1997).
8. A. Robson, *J. Polit. Econ.* **109**, 900 (2001).
9. H. B. Barlow, in *Current Problems in Animal Behavior*, W. H. Thorpe, O. L. Zangwill, Eds. (Cambridge Univ. Press, Cambridge, 1961).
10. P. W. Glimcher, *Decisions, Uncertainty and the Brain: The Science of Neuroeconomics* (MIT Press, Cambridge, MA, 2003).
11. C. S. Sherrington, *The Integrative Action of the Nervous System* (Scribner, New York, 1906).
12. W. T. Newsome, K. H. Britten, J. A. Movshon, *Nature* **341**, 52 (1989).
13. W. T. Newsome, K. H. Britten, C. D. Salzman, J. A. Movshon, *Cold Spring Harbor Symp. Quant. Biol.* **55**, 697 (1990).
14. See also (50) and (51) for a parallel set of studies that presaged and continues to parallel many of these developments.
15. M. N. Shadlen, K. H. Britten, W. T. Newsome, J. A. Movshon, *J. Neurosci.* **16**, 1486 (1996).
16. R. A. Andersen, L. H. Snyder, D. C. Bradley, J. Xing, *Annu. Rev. Neurosci.* **20**, 303 (1997).
17. C. L. Colby, M. E. Goldberg, *Annu. Rev. Neurosci.* **22**, 319 (1999).
18. M. L. Platt, P. W. Glimcher, *Nature* **400**, 233 (1999).
19. B. Knutson, C. M. Adams, G. W. Fong, D. Hommer, *J. Neurosci.* **21**, RC159 (2001).
20. M. P. Paulus, N. Hozack, X. Zauscher, *Neuroimage* **13**, 91 (2001).
21. H. C. Breiter, I. Aharon, D. Kahneman, A. Dale, P. Shizgal, *Neuron* **30**, 619 (2001).
22. D. Kahneman, A. Tversky, *Econometrica* **47**, 263 (1979).
23. J. I. Gold, M. N. Shadlen, *Nature* **404**, 390 (2000).
24. J. I. Gold, M. N. Shadlen, *Trends Cogn. Sci.* **5**, 10 (2001).
25. M. N. Shadlen, K. H. Britten, W. T. Newsome, J. A. Movshon, *J. Neurosci.* **16**, 1486 (1996).
26. M. N. Shadlen, W. T. Newsome, *J. Neurophysiol.* **86**, 1916 (2001).
27. L. P. Sugrue, G. S. Corrado, W. T. Newsome, *Science* **304**, 1782 (2004).
28. P. Janssen, M. N. Shadlen, *Soc. Neurosci. Abstr.* **767**, 2 (2003).
29. B. Coe, K. Tomihara, M. Matsuzawa, O. Hikosaka, *J. Neurosci.* **22**, 5081 (2002).
30. J. V. Von Neumann, O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton Univ. Press, Princeton, NJ, 1944).
31. J. F. Nash, *Proc. Natl. Acad. Sci. U.S.A.* **36**, 48 (1950).
32. J. F. Nash, *Econometrica* **18**, 155 (1950).
33. D. J. Barraclough, M. L. Conroy, D. Lee, *Nat. Neurosci.* **7**, 404 (2004).
34. M. C. Dorris, P. W. Glimcher, *Neuron*, (in press).
35. J. F. Nash, *Ann. Math.* **54**, 286 (1951).
36. P. R. Montague *et al.*, *Neuroimage* **16**, 1159 (2002).
37. D. Ellsberg, *Q. J. Econ.* **75**, 643 (1961).
38. M. Allais, *Econometrica* **21**, 503 (1953).
39. For a review of these issues, see D. Luce, *Utility of Gains and Losses*, vol. 8 of *Scientific Psychology Series* (L. Erlbaum, Mahwah, NJ, 2000).
40. G. Gigerenzer, R. Selten, *Bounded Rationality: The Adaptive Toolbox* (MIT Press, Cambridge, MA, 2002).
41. J. Dickhaut *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3536 (2003).
42. See for example A. Bechara, H. Damasio, D. Tranel, A. Damasio, *Science* **275**, 1293 (1997).
43. J. O'Doherty, M. L. Kringelbach, E. T. Rolls, J. Hornak, C. Andrews, *Nat. Neurosci.* **4**, 95 (2001).
44. A. Rustichini, J. Dickhaut, P. Ghirardato, K. Smith, J. Pardo, in *Games and Economic Behavior*, in press.
45. W. Guth, R. Schmittberger, B. Schwarze, *J. Econ. Behav. Organ.* **3**, 347 (1982).
46. A. Sanfey, J. K. Rilling, J. Aronson, L. E. Nystrom, J. Cohen, *Science* **300**, 1755 (2003).
47. K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11832 (2001).
48. C. D. Frith, U. Frith, *Science* **286**, 1692 (1999).
49. H. L. Gallagher, C. D. Frith, *Trends Cogn. Sci.* **7**, 77 (2003).
50. C. R. Gallistel, *Cognition* **50**, 151 (1994).
51. K. Conover, P. Shizgal, in *Games and Economic Behavior*, in press.

RELEVANT PRIMERS
FROM *CURRENT BIOLOGY*



Primer

Computational theories of vision

Andrew Glennerster

Neuroscientists study the brain at many different levels, from molecular to psychological. Despite progress on many of these levels, there is a disappointing lack of coherence in neuroscience research: we have yet to find an overarching theoretical framework for understanding what the brain does.

Vision is one of the most intensively studied areas of brain function. Yet, even in this field, there are wide disagreements about the goals of cortical processing. In the second half of the twentieth century, there were two important attempts to provide a theoretical framework for understanding vision, by David

Marr and James Gibson. There are now grounds for optimism that these two broad approaches can be brought together to provide a biologically plausible and yet computationally tractable framework for understanding and imitating human vision. I shall focus on attempts to describe vision at the level Marr called 'computational theory'.

Marr

Marr (1982) emphasised that vision was nothing more than an information-processing task. Any such task, he argued, could be described on three levels: (i) computational theory; (ii) specific algorithms; and (iii) physical implementation. The three levels correspond roughly to: (i) defining the problem and setting out how, in principle, it can be solved; (ii) designing a detailed simulation of the process; and (iii) building a working system that will carry it out (see Box 1). The important point is that the levels can be considered independently. As a result, it ought to be possible

to mimic the algorithms underlying biological vision in robots: the only difference would be in how they are implemented physically. This concept of independent levels of explanation remains a mantra of vision research.

Marr attempted to set out a computational theory for vision as a whole. He suggested that visual processing passes through a series of stages, each corresponding to a different representation, from retinal image to '3D model' representation of objects. One problem with this account is that information needs to be passed continually from one coordinate frame to another. There is increasing interest in models that avoid coordinate transformations of this kind, rather using information stored in retinal coordinates for tasks such as object recognition or navigation (for example Mallot, 2000).

Gibson

Like Marr, Gibson had a powerful influence on vision research in the last century. Marr himself wrote

Box 1

Marr's three levels for understanding any visual process, each illustrated using the example of binocular stereopsis.

Level	Definition	Example
Computational theory	Sets out the goal of a process and an outline of how it can be achieved in principle. This includes defining the input, the output and establishing the constraints that will be used in computing one from the other.	For binocular stereopsis, the input is (at least) the left and right eyes' images; the output could be a depth map of features as viewed from a point midway between the two eyes. An example of a constraint is that a point in an image corresponds to one and only one scene point (as the scene point must be opaque for it to be visible).
Algorithm	Shows how a process is to be carried out. Gives details of how the input and output are represented and a set of rules for the transformation between the two.	Stereo algorithms often use sparse sets of features (such as light/dark boundaries) as their input. Coarse-to-fine algorithms compute depth in blurred versions of the input images and use the results to help solve ambiguities at finer scales.
Implementation	Specifies the physical method for carrying out an algorithm, for example in computer hardware or using neurons.	Disparity sensitive neurons have been identified at important stages in the process of stereopsis in animals, some closer to the 'input' stage, others apparently closer to the 'output' stage. However, we do not yet know the algorithm or even the computational theory underlying the process these cells are involved in.

Recent neurophysiological investigations have improved our knowledge of the first stages of processing, in area V1 of the visual cortex, after input from the two retinæ have been combined but before the depth of points has been computed unambiguously (reviewed by Cumming and DeAngelis, 2001). Other studies have identified neurons that appear to be beyond the stage at which depth is computed, because firing of these neurons is remarkably predictive of the perceptual choice an animal will make in a depth-related task (for example, Dodd *et al.*, 2001). There are tantalisingly few synapses between these two neuronal stages, raising the hope that understanding the processes between them may be a tractable problem. The role of these cells in depth perception will only be fully understood, however, when their function can also be described at the levels of an algorithm and a computational theory.

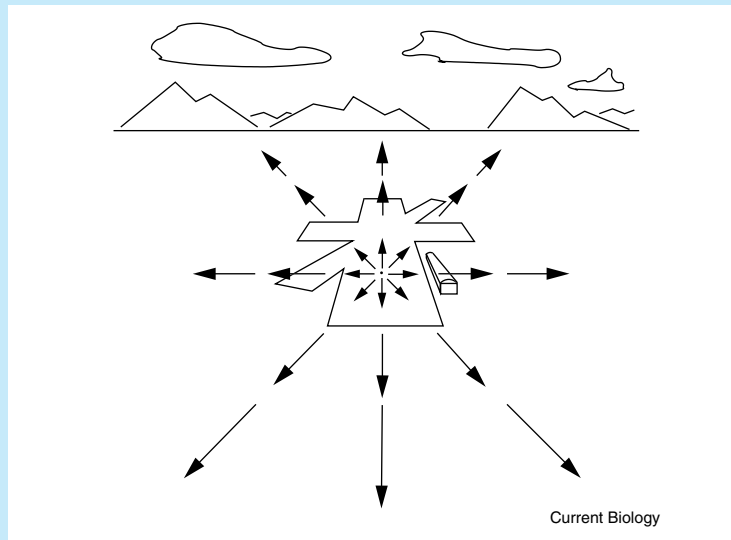
that “in perception, perhaps the nearest anyone came to the level of computational theory was Gibson”. Gibson promoted an ‘ecological’ approach to studying vision, by which he meant that vision should be understood first and foremost as a tool that enables animals to achieve the basic tasks required for life: avoid obstacles, identify food or predators, approach a goal and so on (for example Gibson, 1979).

This viewpoint has gained increasing influence. An emphasis on survival of the organism is a more promising basis for a computational theory of vision than Marr’s assertion that vision is “knowing what is where by looking”. Where Gibson infuriated his contemporaries was in his musings about the brain mechanisms that might generate these ‘ecological’ behaviours (Ullman, 1980). Even when he avoided any mention of brain mechanisms and stuck to what Marr would call the algorithmic level, his proposals were often loose or unclear (see Box 2).

Despite the criticisms, there has been continued interest in exploring the kind of ‘rule-based’ behavioural strategies that Gibson advocated. A recurring theme behind all these strategies is the idea that out of the myriad potential visual signals that could be derived from a moving retinal image, one or two aspects of the information are especially relevant for controlling a particular motor behaviour. Recently, for example, there has been interest in recording the head and eye movements of people carrying out real-world tasks, such as driving. It is clear from these studies how a sequence of simple visuomotor rules or sub-tasks could be linked together to achieve a higher-order goal. Take the task of making a cup of tea, for which Land *et al.* (1999) recorded the entire sequence of head and eye movements. The sub-tasks, such as bringing the hand towards the kettle lid, tend to be straightforward visually guided routines when examined individually. The complexity of behaviours may therefore evolve in two ways: (i) by increasing the

Box 2.

Using optic flow to land a plane.



As the pilot approaches the air field, visible points move outwards, as shown by the arrows. Gibson described this as ‘optic flow’. The point from which the flow emerges indicates the plane’s direction of heading (in computer vision, this is called the ‘epipole’). One strategy for landing the plane is to keep the epipole centred on the runway. (Adapted from Gibson (1979).)

In some ways, the strategy described here is deceptively simple. Gibson said that optic flow was only generated by translation of the observer (movement through space). Rotating the eye does not change the ‘optic array’ and so generates no optic flow. (This is not quite enough, however, to define optic flow: a choice must be made about how to relate different optic arrays. A sensible coordinate frame is one in which distant objects remain stationary, such as the mountains (Glennerster *et al.* 2001)). Retinal flow is more complicated than optic flow. The eyes often rotate with respect to the world as, for example, when you fixate a nearby object and walk past it. The resulting ‘rotational flow’ is added to the ‘pure’ translational flow, which would be generated if you walked past the object fixating a distant point. Thought of in this way, retinal flow is a confusing mixture of two types of signal. Gibson himself was rather unclear on how the visual system was supposed to use ‘optic flow’ given that it must start off with retinal flow.

There are two diverging hypotheses about how the brain deals with retinal flow. One assumes that the visual system extracts the ‘translational’ flow, shown by the arrows in this figure, by subtracting the rotational flow component (reviewed by Lappe *et al.*, 1999). Another assumes that the visuo-motor system uses task-specific strategies that avoid computing translational flow. For example, Cutting *et al.* (1992) suggest that observers could fixate on different points as they walk and, using a simple rule, change their direction of gaze until it is aligned with their direction of heading. The difference between these approaches is at the level of computational theory: they have different goals.

range of different sensory parameters available for controlling the motor system, and (ii) by storing increasingly long sequences of sub-tasks that, when strung together, achieve higher-order goals.

The approaches advocated by Marr and Gibson are not mutually exclusive. Advances in computer vision may help to bring the two together. Mathematical rigour and computational theory are brought to bear here on tasks that, increasingly, must be carried out

in unpredictable, ‘natural’ environments. One current research theme that exemplifies that process of resolution is Bayesian inference.

Bayes

Bayesian inference is sometimes couched in fearsome mathematical terms, but the basic idea is both straightforward and highly relevant to understanding animal behaviour.

The brain receives signals from afferent (sensory) fibres. On the

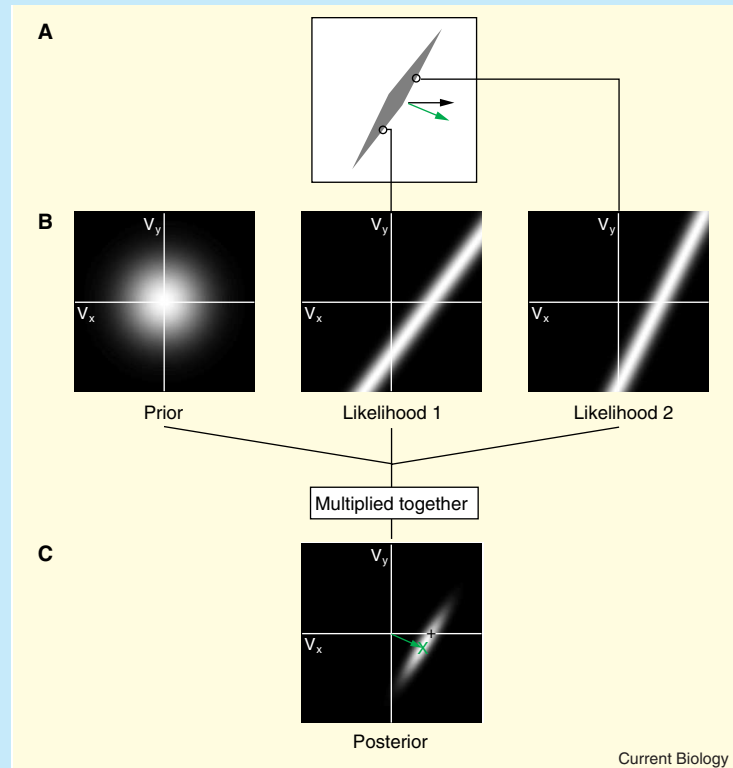
basis of these, and of the information it has stored previously, the brain must generate a response (ultimately, a motor response). A reasonable model for this process is that one response is picked out of a list of possibilities by choosing the most appropriate in the organism's current context (for example, most probably rewarded or least probably punished).

It is here that Bayes' formula is useful. Bayes pointed out that the probability of state *S* being the case (such as 'there is a kettle over to my right') given information *I* (here, the sensory information the brain receives) is directly proportional to two quantities that can, in principle, be estimated in advance, and hence, in the context of the brain, stored in memory. The first quantity is the 'prior' probability of state *S* occurring, *P*(*S*). This makes sense intuitively: if you are forced to guess what the current state of the world is and you have no evidence (or highly inconclusive evidence) at the moment, you should guess a likely rather than an unlikely state (these prior probabilities being determined on the basis of previous experience). For example, if the kettle was on your right the last time you looked, it is a reasonable assumption that the fuzzy grey shape on the periphery of your vision is (still) the kettle.

The second quantity incorporates the actual data, *I*, and gives an indication of how conclusive it is. It is the probability of receiving evidence *I* given that the current state of the world really is *S*, normalised by the total probability of getting information *I* (summed over all possible states). Again, this makes sense intuitively. For example, fuzzy grey shapes are common in peripheral vision and do not always arise from kettles, so the evidence on its own is inconclusive. Fixating the kettle is a good way to improve the evidence. The higher resolution image is richer (and rarer) and more specific to kettles. In general, if sensory input *I* is both rare (*P*(*I*) is low) and also characteristic of state *S* (*P*(*I*|*S*) is high), then information *I* is good evidence that the world is in state *S*. Put more succinctly:

Box 3.

A Bayesian model of a motion illusion.



When a narrow, low-contrast rhombus, as in A, is moved to the right it appears to move down as well (green arrow). This can be understood by (i) considering the set of stimuli that could have produced the edge motion signals the observer receives and (ii) including a 'prior' assumption that objects tend to move slowly (adapted from Weiss *et al.* 2002). The left-hand plot in B shows the assumed 'prior' probability of velocities in horizontal and vertical directions (*V_x*, *V_y*), where intensity is proportional to probability: low velocities are favoured. The centre and right hand plots show the likelihood that an edge moving at each velocity (*V_x*, *V_y*) generated the motion observed at the points marked by the circles. For a high contrast edge, these velocities would all fall along a line. (The fact that the velocity of an edge is not known uniquely is known as the 'aperture problem': it arises because movement in the direction of the edge produces no local motion signal, for example within the circle.) The line is blurred for the low-contrast stimulus because, with some noise in the system, edges moving at other velocities can give rise to the same motion signal.

The plot in C is obtained by multiplying the three plots in B together. This amounts to following Bayes' rule to calculate the probabilities that the real object had a velocity (*V_x*, *V_y*) which gave rise to the two motion signals measured at the circles (known as the 'posterior' probability). (Weiss *et al.* (2002) did this for all points in the stimulus, giving a very similar result.) The mean (or peak) of the distribution is shifted, as shown by the green 'X', away from the true velocity of the rhombus (shown by the '+'). It predicts the direction of motion seen by subjects when presented with this stimulus.

$$P(S|I) = \frac{P(S)P(I|S)}{P(I)}$$

Bayes' rule can be derived from an assertion that the probability of *S* and *I* is equal to that of *I* and *S*. If the two joint probabilities are expressed in terms of conditional events, this becomes *P*(*S*|*I*) *P*(*I*) = *P*(*I*|*S*) *P*(*S*), from which the expression for *P*(*S*|*I*) can be obtained.

Many perceptual phenomena can be explained parsimoniously using a Bayesian approach (see Box 3 and a review by Knill and Richards (1996)).

Bayesian inference fits well with all of Marr's levels of description. It is a useful tool in describing a problem at the level of computational theory, making

explicit what is to be computed and the constraints that are to be used to derive output from input. It can be the basis of specific, working models or algorithms and it can be implemented in a number of ways, such as in neural networks. Following Marr's notion of independence between levels, theories of neural architecture in the brain that might carry out this kind of inference can be developed to deal with the generic quantities $P(S)$, $P(I)$ and $P(I|S)$, without reference to specific stimuli (reviewed by Barlow, 2001).

At the same time, Bayesian approaches fit well into the evolutionary or ecological perspective that Gibson advocated. A simple organism, with a simple behavioural repertoire, needs only to divide information about the organism's state with respect to the world into a small number of categories. It can use its motor system to move between these categories (this is, after all, the only way it can know that a motor movement has been successful). A more complex behavioural repertoire requires a greater number of states to be discriminated reliably. This means that sensory systems must evolve to help an organism discriminate between the contexts in which it will generate different motor outputs. At various stages in a task, the sensory parameters that are most helpful in discriminating (and hence controlling) movements will be quite different (see section on Gibson). This leads to a view of the cortex as a pool from which evidence can be drawn. From moment to moment, the neurons with the most relevant information may be located in quite different parts of the cortex, according to the demands of the task.

Time

Proposals about how neural signals are combined to give rise to visual percepts include some that do not provide a description of the solution at the level of a computational theory: an example is the idea that

synchronised oscillations in the firing rate of neurons in different parts of the brain could account for perceptual phenomena (Singer, 1998). As they stand, such theories have little explanatory power.

A general weakness of these and many other theories is the failure to consider how representations could be built up over time. There is a tendency to assume that neural responses could somehow be combined to generate a vivid reconstruction of the scene in an instant. Not only is this a daunting prospect, but the computational problem is amplified as time is brought into the equation. Having reconstructed the world, heaven forbid that the observer now move their eyes or their head! That would entail a new reconstruction and a new problem of relating it to the one created a moment ago. More promising approaches focus on the small, discrete goals of one epoch (for example, a period of fixation) and how these could be combined, like the pieces of a jigsaw, into a richer representation (for example Ballard *et al.* 1997, Land *et al.* 1999, Rensink, 2002).

This view brings Marr and Gibson's ideas together in another way. Gibson emphasised the role of vision as a tool for action. One of the things that makes human vision special is our ability to carry out tasks involving long sequences of movements, each one simple if considered in isolation, to achieve our goals. The processes involved in building up a vivid, detailed visual representation are perhaps best seen as a by-product of that ability, taking time and being divisible into purposeful steps. But as Marr emphasised, whatever the processes turn out to be, emulating them computationally is the best way to understand them fully.

Acknowledgements

Supported by the Royal Society. I am grateful for comments from Bruce Cumming, Tam Curnow, Miles Hansard and Anya Hurlbert.

References

- Ballard, D.H., Hayhoe, M.M., Pook, P.K. and Rao, R.P.N. (1997). Deictic codes for the embodiment of cognition. *Behav. Brain Sci.* 20, 723–742.
- Barlow, H.B. (2001). Redundancy reduction revisited. *Netw. Comput. Neural Syst.* 12, 241–253.
- Cumming, B.G., and DeAngelis, G.C. (2001). The physiology of stereopsis. *Annu. Rev. Neurosci.* 24, 203–238.
- Cutting, J.E., Springer, K., Braren, P.A. and Johnson, S.H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *J. Exp. Psychol. Gen.* 121, 41–72.
- Dodd, J.V., Krug, K., Cumming, B.G. and Parker, A.J. (2001). Perceptually bi-stable three-dimensional figures evoke high choice probabilities in cortical area MT. *J. Neurosci.* 21, 4809–4821.
- Gibson, J.J. (1979). *The ecological approach to visual perception* (Houghton Mifflin).
- Glennester, A., Hansard, M.E. and Fitzgibbon, A.W. (2001). Fixation could simplify, not complicate, the interpretation of retinal flow. *Vision Res.* 41, 815–834.
- Knill, D. and Richards, W. (1996). *Perception as Bayesian Inference* (Cambridge University Press).
- Land, M.F., Mennie, N. and Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception* 28, 1311–1328.
- Lappe, M., Bremmer, F. and van den Berg, A.V. (1999). Perception of self-motion from visual flow. *Trends Cogn. Sci.* 3, 329–336.
- Mallot, H.A. (2000). *Computational vision: information processing in perception and visual behaviour*, 2nd edn (MIT Press).
- Marr, D. (1982). *Vision: a computational investigation into the human representation and processing of visual information* (W.H. Freeman and company).
- Rensink, R.A. (2002). Change detection. *Annu. Rev. Psychol.* 53, 245–277.
- Singer, W. (1998). Consciousness and the structure of neuronal representations. *Philos. Trans. R Soc. Lond. B. Biol. Sci.* 353, 1829–1840.
- Ullman, S. (1980). Against direct perception. *Behav. Brain Sci.* 3, 373–415.
- Weiss, Y., Simoncelli, E.P. and Adelson, E.H. (2002). Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604.

University Laboratory of Physiology,
Parks Road, Oxford, OX1 3PT, UK.
E-mail: ag@physiol.ox.ac.uk

Primer

Decision making

Jeffrey D. Schall

Daily we encounter numerous alternative courses of action among which we must choose. These alternatives vary in complexity and consequence. Some alternatives are considered lightly, such as the way to turn on the journey home. Others require fervent deliberation, such as Hamlet's agonizing decision whether to end his life. The neural basis of decision making has been reviewed extensively [1–4], so the goal of this primer is to orient the interested reader to this growing literature with emphasis on neurophysiological work, rather than the expansive literature on functional brain imaging and neuropsychology. Another goal is to articulate the key concepts of choosing, deciding, intending and acting [5,6]. Dwelling on terminology may seem an unnecessary tangent, but science travels on its vocabulary — inconsistent and vague terms can only yield confusion. This is all the more important when the object of this scientific investigation ultimately is nothing less than human agency.

Choice

A choice is required when an organism is confronted with alternatives for which an action is required to acquire or avoid one of the alternatives because of a desire, goal or preference. In its most fundamental sense, a choice is an overt action performed in the context of alternatives for which explanations in terms of purposes can be given. Consider the options of selecting between two envelopes, A and B, containing different amounts of money (Table 1). Choice #1 is easy because it offers the guarantee of obtaining \$1,000 by selecting envelope A, instead of \$1 by selecting envelope B. Choice #2 requires more thought because the amount of money is unpredictable, but a suitable analysis leads to a wise choice.

Choice #3 requires deliberation to contend with a paradox of choice prediction [7]. Regardless of the complexity of the alternatives, though, you must execute your choice through some action, such as grasping one of the envelopes.

Choices have particular characteristics. First, choices are evaluated as good or bad according to whether goals are achieved and consequences are as expected. Second, choices take time; a choice process evolves from a state of more or less equipotentiality immediately after the alternatives are presented to a state of commitment before the overt action is performed. Third, with prior knowledge of the alternatives and preferences, some choices can be predicted.

According to this definition, neurophysiological correlates of choice behavior are studied whenever an experimental subject has alternative responses. This has been studied, for example, with psychophysical discrimination [8] and visual search [9]. With easily distinguished alternatives and predictable reinforcements, like choice #1, the neural representation of the stimuli is resolved unambiguously, leading to an earlier and more accurate motor response. When the stimuli are more difficult to distinguish [10,11] or the reinforcement becomes less predictable [12,13], like choice #2, the neural representation of the alternative choices takes more time to resolve, and that representation can be modulated by the expected value of the choice.

Finally, although no neurophysiological studies have been conducted in the context of Newcomb's paradox, new studies are exploring situations in which the experimental subject chooses alternatives supposedly based on the actions of another agent [14]. Choices made in the context of the choices of another agent are also guided by reinforcement history, but also with some sensitivity to the strategies employed by the other agent. This progression of experimental conditions should make clear that some choices are easy and others are hard. When alternatives are vague, payoffs are unclear and stakes are high, then it seems natural to regard the subjects as making a decision before they act.

Decision

The term decision is used casually and technically in several, not entirely consistent senses. On the one hand, decision theory describes the characteristics of effective behavior in relation to expected consequences [15]. For example, the best solution to the second choice in Table 1 is to select the envelope for which the product of the probability and the monetary amount is greatest. This is the domain of decision theory and related frameworks, such as economics and game theory. A very active area of inquiry of late has employed concepts of decision theory and related disciplines in the investigation of the neural preliminaries of choice behavior [16].

Table 1. Choices based on predicted monetary reward.

	Envelope A		Envelope B	
	Condition	Payoff	Condition	Payoff
#1	100%	\$1,000	100%	\$1
#2	9%	\$10,000	90%	\$1,000
#3	100%	\$1,000	If only B	\$1,000,000
			If A and B	\$0

Two envelopes labeled A and B are offered for you to obtain. Choice #1 is between two transparent envelopes; envelope A contains \$1,000 and envelope B, \$1. Choice #2 is between two opaque envelopes. Envelope A has a 9% chance of containing \$10,000 and envelope B has a 90% chance of containing \$1,000. Choice #3 is Newcomb's paradox. Envelope A is transparent and contains \$1,000. Envelope B is opaque and will contain either \$1 million or nothing. You will be given the choice of taking either B alone or A with B. Before you are presented with the envelopes, an infallible brain scan is done to predict your choice. If the prediction is that you will choose only envelope B, then \$1 million is placed in it. If it is predicted that you will choose both envelope A and B, then nothing is put in envelope B.

On the other hand, we can refer to a decision as a process that results in the overt act of choosing. Such a process must have a particular architecture, but measures of outcome derived from decision theory do not uniquely specify mechanism. Decision as a process has two logically and practically distinct meanings. We must distinguish between *decide to*, which is a selection between alternative actions, and *decide that*, which is a selection between alternative categories of an image or concept. The logical distinction is easy to see. You can *decide that* falsely, but you cannot *decide to* falsely — it is not intelligible. The *decision that* envelope B always contained \$1 million is false, but in what sense can you *decide* falsely to take envelope B? You just take it or you do not. *Decide to*, like choosing, is judged as good or bad, but not as true or false.

The fact that choices can be predicted makes it possible to choose in advance: it is intelligible to say, “I will choose envelope B when given the opportunity.” This sense of choice is more related to decision. But important distinctions must be recognized between ‘choose’ and ‘decide to’. Whereas ‘choice’ refers most directly to the final commitment to one among the alternative actions, ‘decision’ refers most directly to the deliberation preceding the action. Deliberation involves weighing the options which amounts to contrasting the characteristics of the alternatives in working memory. Such deliberation is rarely (if ever) only rational, for emotional valence influences decisions too [17].

The polarity of deciding and choosing can be highlighted by contrasting the three choices in Table 1. For any of the three, you must enact a choice by grasping an envelope (you could just point or speak to choose through symbolic actions, but some movement must occur). But it is likely that, before you can make choices #2 or #3, you must invest more effort to comprehend the alternatives, understand the differences between them and identify which would be most satisfying. The course of this

deliberation reveals another defining feature of decisions that distinguishes them from choices: decisions cannot be predicted, even by the agent. If you can say what you will decide, then you will have decided.

In other words, decisions, like perceptions, just happen; introspection cannot find the source of a decision [18]. By monitoring the activity of the neurons that produce the decision as described above, however, it becomes possible to predict the choice. Moreover, it is also possible to manipulate decisions by electrical stimulation [19]. But this ability to predict and manipulate choices has strict limits; most fundamentally, it requires proper experimental control of numerous extraneous variables. Thus, while it is appropriate to contemplate the ethical and legal ramifications of this technical breakthrough [20], the social fabric is unlikely to be torn by it very soon.

Choice #3 provides a useful domain in which to clarify further these ideas. It is based on the premise that a person’s choice can be predicted reliably (by a brain scan perhaps). The rules of this choice are explained in Table 1. Imagine that you have observed others make this choice and that every time they took only envelope B, it contained \$1 million; and every time they took both, envelope B was empty. Would you reason that the choice prediction is reliable and so choose only envelope B to become a millionaire? Or would you reason that, once the prediction was made and you are presented with the selection, the contents of the envelopes will not change, and so choose both envelopes so as not to forego the obvious \$1,000? Both arguments seem sound, but they cannot both be correct, so which is wrong? Your reaction to the contents of the envelopes opened after making choice 3 may be pleasure or distress, according to whether the outcome was consistent with your intention. What did you mean to do?

Intention

The term intention is complex [21]. The disposition to perform some

act is a central feature of an intention, but intention cannot be identified entirely with response preparation. A statement of intention must also answer “why was that done?”. Of course, one answer might be the causal path through neurons to muscles, but this is incomplete. A satisfactory explanation must address the *reasons* for the action based on preferences, goals and beliefs. In other words, to judge whether a movement was intended, one must refer to the agent’s beliefs about which action must be performed under what circumstances to bring about the desired object of the intention. A consequence of this is that intentions may not be realized; success can be judged only with reference to the description of the goal and the conditions under which it could be achieved [22]. Furthermore, a particular movement may be intentional under one description but not under another; for example, an eye may wink or blink. The neurophysiology literature has used the term intention to refer only to response preparation [23]. Until nonhuman primates can report what they meant to do, the scientific study of intentionality may be restricted to humans [24].

Action

The term action has been used above, but it is a complex concept [25]. Often one does one thing (obtain money) by doing something else (pick up an envelope). But all such descriptions of actions reduce to a *basic action* that we perform without any preliminaries, such as moving a hand (to grasp an envelope to obtain money). We cannot say how we move our hand; it just happens when we will it. The neurophysiological basis of producing basic actions such as gaze shifts and reaching are reasonably well understood [26].

Defining a body movement as an action depends on context. A purposeful action (a wink) is distinguished from a mere event (a blink) by reference to some intelligible plan, because actions are performed to achieve a goal. In other words, actions have *reasons* (“I did it for...”), but events just have *causes* (“It happened

because..."). Reasons for actions are explanations in terms of purposes, that is, intentions. Thus, a particular movement may be intentional under one description but not under another. With human subjects it is possible to distinguish the brain states producing intentional versus unintentional movements [27,28].

But if all actions are caused by neurons firing and muscles contracting, how can there be any reasons for actions? Evidence for many-to-one mapping of brain states onto movements has important implications for elucidating a neural basis of intentional action [29]. If the mapping of neural activity onto movement were one-to-one, the causal basis of movements would be clear: a particular action follows necessarily from a given brain state as reliably as a reflex. While such an automatic causal process seems an adequate account of certain kinds of movements (such as blinks), it cannot provide a satisfactory account of other kinds (such as winks). Intended movements are owned ("I did") while unintended movements are not ("it happened"). In other words, we distinguish the *cause of* from the *reason for* movements [30].

In fact, some have argued that a many-to-one mapping of neural activity onto cognition and behavior provides room for intentional reasons within neural causes [31]. If a given basic action can arise from different brain states, then the dependence of the behavior on an intention can hold in virtue of the content of the representation of the intention and not its neural realization, that is, the content which answers "why did you do that?". Thus, a movement can be called an intentional action if and only if it originates from a cognitive state with meaningful content, and this content defines the cognitive state's causal influence.

This analysis depends on whether the brain knows what it means to do. Recent research has shown that the medial frontal lobe registers error and success [32]. For example, in monkeys performing a task requiring the interruption of a planned action,

neurons signaling errors and omission of earned reinforcement were observed in the medial frontal lobe [33]. Such signals can be used to adjust behavior and provide the basis for distinguishing "I did" from "it happened".

References

- Schall, J.D. (2001). Neural basis of deciding, choosing and acting. *Nat. Rev. Neurosci.* 2, 33–42.
- Glimcher, P.W. (2001). Making choices: the neurophysiology of visual-saccadic decision making. *Trends Neurosci.* 24, 654–659.
- Romo, R., and Salinas, E. (2001). Touch and go: decision-making mechanisms in somatosensation. *Annu. Rev. Neurosci.* 24, 107–137.
- Gold, J.I., and Shadlen, M.N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends Cogn. Sci.* 5, 10–16.
- Nowell-Smith, P.H. (1958). Choosing, deciding and doing. *Analysis* 18, 63–69.
- Evans, L.L. (1955). Choice. *Philosophy Qtr.* 5, 303–315.
- Nozick, R. (1969). Newcomb's problem and two principles of choice. In *Essays in Honor of Carl G. Hempel and N. Rescher*, eds. (Holland, Dordrecht: D. Reidel).
- Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu. Rev. Neurosci.* 21, 227–277.
- Schall, J.D. (2002). The neural selection and control of saccades by the frontal eye field. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, pp1073–1082.
- Sato, T., and Murthy, A. (2001). K.G., Thompson, J.D. Schall. Search efficiency but not response interference affects visual selection in frontal eye field. *Neuron* 30, 583–591.
- Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22, 9475–9489.
- Kawagoe, R., Takikawa, Y., and Hikosaka, O. (2004). Reward-predicting activity of dopamine and caudate neurons—a possible mechanism of motivational control of saccadic eye movement. *J. Neurophysiol.* 91, 1013–1024.
- Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science* 304, 1782–1787.
- Barracough, D.J., Conroy, M.L., and Lee, D. (2004). Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* 7, 404–410.
- Pratt, J.W., Raiffa, H., and Schlaifer, R. (1995). *Introduction to Statistical Decision Theory*. (Cambridge, Mass.: MIT Press).
- Glimcher, P.W., and Rustichini, A. (2004). *Neuroeconomics: The consilience of brain and decision*. *Science* 306, 447–452.
- Bechara, A., and Damasio, H. (2000). A.R. Damasio. Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex* 10, 295–307.
- Dennett, D.C. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. (Cambridge, Mass.: MIT Press).
- Cohen, M.R., and Newsome, W.T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.* 14, 169–177.
- Farah, M.J. (2002). Emerging ethical issues in neuroscience. *Nat. Neurosci.* 5, 1123–1129.
- Aune, B. (1967). Intention. In *The Encyclopedia of Philosophy*, Volume 4 (New York: Macmillan).
- Heckhausen, H., and Beckmann, J. (1990). Intentional action and action slips. *Psychol. Rev.* 97, 36–48.
- Snyder, L.H., Batista, A.P., and Andersen, R.A. (2000). Intention-related activity in the posterior parietal cortex: a review. *Vision Res.* 40, 1433–1441.
- Haggard, P., and Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Conscious. Cogn.* 12, 695–707.
- A.I. Goldman (1970). *A Theory of Human Action*. (Englewood Cliffs, N.J.: Prentice-Hall).
- Schall, J.D., Hanes, D.P., and Taylor, T.L. (2000). Neural control of behavior: countermmanding eye movements. *Psychol. Res.* 63, 299–307.
- Keller I, Heckhausen H. (1990). Readiness potentials preceding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalogr. Clin. Neurophysiol.* 76, 351–361.
- Kato, M., and Miyauchi, S. (2003). Functional MRI of brain activation evoked by intentional eye blinking. *Neuroimage* 18, 749–759.
- Schall J.D. (2004). On building a bridge between brain and behavior. *Annu. Rev. Psychol.* 55, 23–50.
- Davidson, D. (1963). Actions, reasons and causes. *J. Philos.* 60, 685–700.
- Juarrero, A. (1999) *Dynamics in Action: Intentional Behavior as a Complex System*. (Cambridge, Mass.: MIT Press.)
- Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science* 306, 443–447.
- Ito, S., Stuphorn, V., Brown, J.W., and Schall, J.D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermmanding. *Science* 302, 120–122.

Center for Integrative & Cognitive Neuroscience, Vanderbilt Vision Research Center, Vanderbilt University, Nashville, Tennessee 37203, USA.

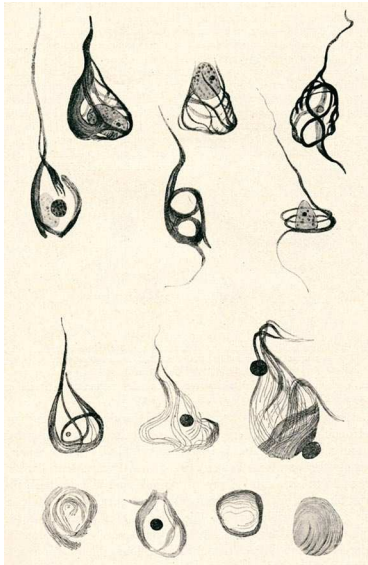


Figure 3. Drawings of neurofibrillary tangles by Alzheimer and published in his 1911 paper.

more than an outstanding scientist who contributed greatly to our knowledge of the histopathology of various brain diseases. He was “first and foremost a psychiatrist who strove to advance psychiatry by using a microscope”. Contrary to the growing movement that regarded disabled people as inferior, Alzheimer treated his patients with great compassion.

Today, Alzheimer's name is associated with one of the cruellest diseases and the mere mention of his name conjures up associations of inexorable mental decline. However, it was his genuine interest in the troubles of his patients and his discovery of the pathological basis of the disease that paved the way to a better understanding of the processes underlying Alzheimer's disease. These fundamental discoveries will no doubt contribute to any future treatments for the affliction. A century after Alzheimer's first description of the disease, the growing number of elderly individuals worldwide makes the need for such treatments increasingly pressing.

Medical University of Vienna, Center for Brain Research, Spitalgasse 4, A-1090 Vienna, Austria.
E-mail: ralf.dahm@meduniwien.ac.at

Primer

Evolution of the neocortex

Jon H. Kaas

Humans have exceptional skills and abilities that largely depend on their brains. One of the major questions of modern neuroscience is to understand how the human brain evolved. In general, we share brain features and functions with other animals in proportion to how closely we are related to them. Thus, we are more similar to our closest living relatives, chimpanzees and bonobos, than to more distant relatives, such as rats and mice, or even more distant still, zebrafish. As we make inferences about the evolution of the human brain, or any other complex brain, such as elephant or whale brains, from studies of other animals, we benefit from an increasing understanding of phylogeny and thus our relationships to other animals.

Tracing the course of human brain evolution

The first question in tracing the course of human brain evolution is how far back to start. We could start with the first nervous systems that emerged with the evolution of multicellular life forms, or the early bilateral nervous system that insects and vertebrates share, but that inverted with respect to the dorsoventral body axis in vertebrates compared to insects. However, for the present discussion, it is practical to start with early mammals, and focus on the neocortex, a part of the brain that was greatly modified from the thin dorsal cortex of reptiles to form the thicker, layered neocortex of mammals. As a result of over 200 million years of divergent evolution, the over 4600 species of extant mammals now vary greatly in overall brain size, and especially the amount of neocortex. The human brain in particular is dominated by two very large, highly convoluted

sheets of neocortex, the most enlarged part of the human brain.

There are at least four complementary approaches we can use to determine the course of forebrain and neocortex evolution from early mammals to humans. One is to compare the brains of extant mammals and assume that features or characteristics held in common have been retained from a common ancestor. Today, this comparative approach usually involves a cladistic character analysis, and several formal procedures have been described. In brief, any group of mammals that have descended from a common ancestor forms a clade, and features (characters) held in common are parsimoniously attributed to that common ancestor. Thus, if we consider the clade of humans and our closest living relatives, the bonobos and chimpanzees, then brain features we share are likely to have been retained from the common ancestor of humans, bonobos and chimpanzees. If we consider the clade that includes all mammals, then features widely present in mammals across the major branches of the mammalian radiation are likely to have been retained from the first mammals. The usefulness of this approach depends on how much we know about the brains of extant mammals.

A second approach is to study the fossil record. Because the soft tissue of brains does not fossilize, this record does not tell us anything about the internal organization of brains. However, skulls do fossilize, and the internal brain case of the skulls of mammals conforms closely to the size and shape of the brain. Thus, endocasts of the brain cases can reveal the size, shape, and even folding patterns in the cortex of the brains of long extinct mammals. Brain size — especially in relation to body size — is an important guide to brain computing power and intelligence, while differences in brain shape suggest differences in the sizes of functional compartments of the brain. Finally, the pattern of cerebral fissures apparent in the endocasts can suggest the locations of functional divisions of cortex.

A third approach that seems very promising is to make conclusions and predictions based on scaling rules for brain size. As the functions of neurons are greatly influenced by brain size, larger brains are not simply larger versions of small brains. Most notably, larger brains have more neurons rather than just larger neurons. This increase in cell number makes it more difficult for each neuron to maintain its connections with the same proportion of other neurons in the brain. And as the brain gets bigger, interacting neurons get further apart, requiring longer axonal connections. As distance equals time in the nervous system, longer axons need to be thicker to maintain short conduction times. Longer, thicker axons take up more space. To reduce such connection problems, large brains are likely to be more modular in organization, emphasizing local as opposed to long-distance connections. For example, large brains might have more cortical areas, not just larger cortical areas. In addition, cortical areas in each cerebral hemisphere might specialize in different functions, reducing the need for large numbers of thick, fast communicating axons between hemispheres.

Other scaling rules have been empirically derived, such as the 'late makes great' rule. In a range of mammals, Barbara Finley and co-workers compared the size of major subdivisions of the brains to the total brain size, thereby providing evidence that brain parts made late in the course of brain development are disproportionately enlarged in bigger brains. To a great extent, this means that mammals with the largest brains devote proportionally more of their brains to neocortex than mammals with the smallest brains. This, of course, does not mean that neocortex is organized in the same way in all mammals with large brains, but it does provide a useful prediction about the proportions of brain parts in relation to overall brain size.

A fourth source of information that can usefully guide and inform theories of brain evolution comes

from our growing understanding of brain development. Studies of the expression, function, and regulation of the genes involved in the patterning of nervous systems help us understand the constraints and opportunities imposed on the genetic toolbox that both preserve ancestral features and allow for endless variation in brain organization and function.

Given the burgeoning information generated by these four approaches, what do we know about the evolution of the human brain from that of the first mammals? The short answer is far too much to tell here, but a short outline is possible. Let's start over 200 million years ago with the neocortex of the first mammals.

The neocortex of early mammals

The fossil record tells us that early mammals were typically small, mouse- to cat-sized and that they had small brains with very little neocortex. We know that olfaction was important, because a large proportion of the forebrain was made up of the olfactory bulb and olfactory (piriform) cortex. In contrast, the neocortex was small, and confined to a thin cap on the top of the forebrain that did not extend over the midbrain, nor have any fissures. It is hard to find any extant mammals today that have so little neocortex. This tells us that the neocortex is generally useful, and that more is generally better. However, there are developmental and metabolic costs to having more neocortex, and so some mammals have found ways to persist and maintain sufficient reproductive success with very little neocortex.

Tenrecs of Madagascar (Figure 1A) are small insect-eating mammals, much like the first mammals. Tenrecs are now recognized as members of the great superorder, Afrotheria, rather than the Insectivore order with moles, shrews, and hedgehogs in the superorder, Laurasiatheria. In shape and size, the forebrain of tenrecs closely resembles that of early mammals. The neocortex of tenrecs is small, thin, and poorly differentiated into layers, cell types, and architectonically distinct regions compared to most other

mammals. In addition, the cortex is divided into few functional areas (Figure 1B). There is evidence for a primary somatosensory area, at least one additional somatosensory field, one or more auditory fields, primary and secondary visual fields, and a primary motor field. There are also limbic and retrosplenial fields in the cortex of the medial wall of the cerebral hemispheres, orbital and frontal fields ahead of motor cortex, one or two fields in the temporal cortex, and perirhinal cortex. But overall, the number of cortical fields is small, on the order of 15–20 fields, compared to an estimated 150 or more fields in the neocortex of humans. When small-brained members of the other major branches of the mammalian radiation are considered — such as opossums, armadillos, rats and hedgehogs — the same pattern of cortical organization emerges. The neocortex is small, the architectonic differentiation of cortex is limited and the number of proposed cortical areas is in the range of 15–25.

Consistently, primary (S1) and secondary (S2) somatosensory areas, primary (V1) and secondary (V2) visual areas as well as one or more auditory areas are found, all of which occupy much of the cortical space. These sensory areas are present in all or nearly all mammals investigated, pointing to an ancient origin. Orbital frontal and limbic systems also appear to have been in place. A primary motor area and usually one or more secondary motor areas are found in all placental (eutherian) mammals studied, but not in marsupials, and not with any certainty in monotremes. Overall, it appears that the neocortex of early mammals was small and poorly differentiated into structurally different areas with different cell types. The cortex was dominated by sensory areas, including the well-known primary and secondary fields, but a separate motor area likely did not emerge until the advent of placental mammals. All these cortical changes represent a great advance over the rather simple dorsal cortex of the reptilian ancestors of mammals.

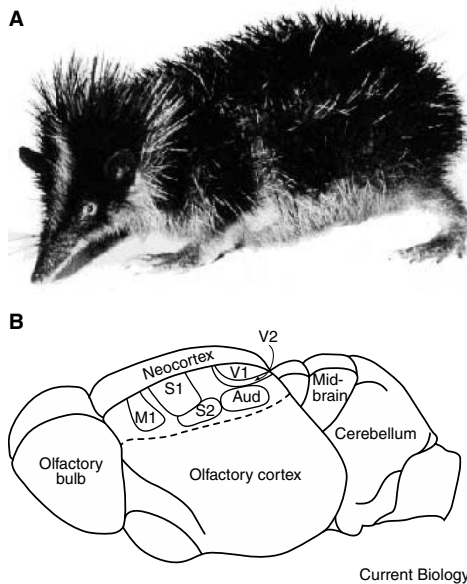


Figure 1. The tenrec brain as a model for the brains of early mammals.

(A) A small Afrotherian mammal, the Madagascar tenrec (*Echinops telfairi*) sharing many characteristics with the earliest mammals. (B) A dorsolateral view of the tenrec brain with major parts and subdivisions of neocortex. Neocortex is small and subdivided into few cortical areas, mainly primary and secondary visual (V1 and V2) and somatosensory (S1 and S2) areas, a primary motor area (M1) and an auditory field (Aud).

The neocortex of early primates

Most early primates were small, nocturnal and adapted to the fine branches of bushes and trees, where they fed on fruits, leaves, insects and small vertebrates. This lifestyle depended on improved visual and sensorimotor abilities. The fossil record indicates that early primates had a considerably larger neocortex than other early mammals, but not as much as most extant primates. In particular, the temporal cortex was expanded. This area was most likely devoted to vision, as it serves this purpose in extant primates and the nearest relatives of primates, tree shrews.

Early primates gave rise to the three major groups of extant prosimians (galagos, lorises and lemurs) and to anthropoid primates, including by most estimates the highly specialized tarsiers, as well as monkeys, apes and humans. Some of the present-day prosimians, such as galagos (Figure 2), resemble early primates in brain size and shape. While primate brain organization has been experimentally studied mainly on monkeys, especially the large Old World macaque monkeys, studies on galagos have been very important in revealing brain features that are common to both prosimian and anthropoid branches of primate evolution, and thus are likely to have been present in early primates (Figure 2).

Given the expanded temporal lobes of early primates, it is not surprising that all studied extant primates demonstrate a greatly modified visual system, with an increased number of visual cortical areas. The first (V1) and second (V2) visual areas are enlarged and specialized for detailed vision, while a number of other visual areas are partially segregated into a dorsal stream of visual processing more devoted to action and a ventral stream more devoted to object identification. The upper temporal lobe of all studied primates contains at least four areas in the dorsal stream that have not been identified in any non-primate mammal, the middle temporal visual area and three associated areas. The ventral stream includes the dorsolateral visual area (V4) and several subdivisions of the ventral temporal lobe. The dorsal stream areas project to an expanded region of the posterior parietal cortex, where a series of areas use visual, somatosensory and auditory inputs to guide reaching and other behaviors via connections with motor and premotor areas of frontal cortex. The ventral stream areas relay information to the frontal cortex where it is stored in short-term memory to guide behavior. These shared features of primate brains, together with shared alterations in the organization of

the visual thalamus and midbrain, indicate that major changes in the visual system had already occurred by the time of the early primates. These changes in brain organization improved visually guided hand and body movements, and the visual detection and identification of food items, predators, and conspecifics.

The other remarkable change in neocortical organization of early primates was the elaboration of the motor and premotor cortex. The mammalian relatives of primates — such as tree shrews and rats — have a primary motor field, M1, and a premotor field, M2. Neither field is highly devoted to forelimb movements. In contrast, both prosimian and anthropoid primates have a large, well-differentiated M1 with a middle portion where electrical stimulation easily evokes forelimb and hand movements. In addition, there are dorsal, ventral and supplementary premotor areas, as well as several cingulate motor areas. Thus, early primates emerged with major changes in the organization of the visual and motor systems, especially at the cortical level. While other innovations in cortical organization also occurred, including those in the somatosensory and auditory cortex, they were not as impressive as the alterations of the visual and motor cortex.

Advanced Old World monkeys and apes

Early anthropoids gave rise to tarsiers, New World monkeys and Old World monkeys. We do not know much about brain organization in tarsiers, except that they do have extremely large eyes and a very specialized visual system. New and Old World monkeys share many features of cortical organization while differing in several ways. In the well-studied brains of Old World macaques, we find an elaboration of the somatosensory cortex, so that four well-differentiated areas in the anterior parietal cortex, areas 3a, 3b, 1 and 2 of Brodmann, contain orderly representations of the contralateral body, as they do in humans. In addition, the number

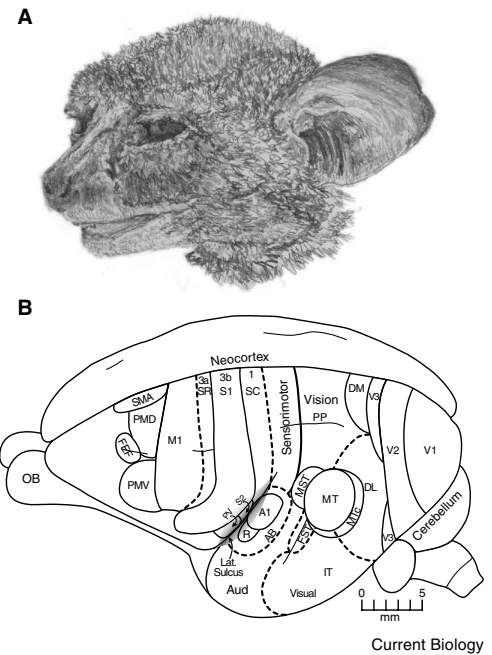
of higher-order somatosensory representations in the cortex of the lateral sulcus has increased, and serial processing has become emphasized so that at least two of these areas, S2 and the parietal ventral area, depend on input from the anterior parietal cortex for activation, rather than on direct thalamic inputs as in prosimians. The visually and somatosensory dominated portions of the posterior parietal cortex have expanded, the number of visual areas has increased, subdivisions of premotor cortex have been added, and the dorsolateral prefrontal cortex has expanded, along with many other modifications. Most of these changes likely were present in the monkeys that gave rise to apes.

The apes that gave rise to chimps, bonobos and humans

Humans have diverged so far in brain size, body form and behavior from the common ancestor of chimpanzees and humans that chimpanzee brains are often considered to be reasonable representations of the brain of the common ancestor. For this reason, it is especially important to learn all we can about the brains of chimpanzees and bonobos and, for comparison, other apes. Because such research falls under restrictions that are similar to those for research on humans, current studies largely involve comparisons of the histological structure of cortical areas and other parts of the brains of apes and humans. These studies hold much promise, as for example, the structural organization of even the primary visual cortex is considerably different in chimpanzees and humans. These differences suggest that humans have altered the distribution of visual information from the primary visual cortex to other visual areas. Possibly other procedures, such as functional magnetic imaging, will be adapted to explore the functional organization of chimpanzee brains for further comparison to humans. It is also clear that chimpanzees have asymmetries in the temporal lobe that are seen to a greater extent in human brains. Thus,

Figure 2. The galago brain.

(A) The cat sized prosimian primate, *Galago garnettii* (also known as *Otolemur garnettii*). The brain of this prosimian may help us understand the evolution of primate brains. (B) A dorsolateral view of a galago brain. The somatosensory areas include area 3b, area 1 or the caudal somatosensory area (SC), area 3a or the rostral somatosensory area (SR), the second area (S2), the parietal ventral area (PV) and other areas buried in the lateral sulcus. Visual areas include the first (V1) second (V2) and third (V3) visual areas, the dorsomedial visual area (DM), the dorsolateral visual area (DL), the middle temporal visual area (MT), the medial superior temporal area (MST), and the fundal area of the superior temporal sulcus (FST).



Current Biology

these asymmetries first appeared in our ancestors for reasons other than mediating language. Our understanding of the huge gap between humans and monkeys in brain complexity and function can be filled only through comparative studies in apes.

Modern humans and their close fossil relatives

The hominin clade includes modern humans, who emerged about 190 thousand years ago, and all extinct species more closely related to humans than our closest living relatives, the chimpanzees. The hominin clade diverged from the one leading to chimps 5–8 million years ago, with *Australopithecus* recognized as a very early member of the clade. Subsequent hominins include *Homo habilis* and *Homo erectus*. Given that all our hominin relatives are extinct, and our understanding of the functional organization of the chimp brain is limited, what can we say about the last 5–8 million years of hominin brain evolution? Here we are limited to the fossil record, which is steadily improving. The most important observation is that brains increased greatly in size, especially over the last 1.5 million years, from sizes similar to those of the great apes to modern human brains three times as big. There are a few other hints from the

fossil record. For example, there is evidence from the position of a fissure in the endocrasts that the reduction in the proportional size of the primary visual cortex of humans compared to chimps occurred early in the hominin line. There is also evidence that hemispheric asymmetries associated with right-handedness and language increased from early to later ancestors. But the increase in brain size remains the clearest finding. From this important observation, we can make several tentative conclusions, based on probable solutions to scaling problems inherent in such an increase in brain size. First, it is likely that the largest of cortical fields stabilized in size, and did not proportionally increase with the brain. As cortical areas increase in size, communication between neurons within an area becomes increasingly longer and difficult to maintain. The functions of an area change as the area becomes less integrative. The best evidence that some stabilization of areas in size actually occurred is that the easy-to-measure primary visual cortex (area 17) is about the same size in chimpanzee and human brains. Second, the brain should become much more modular in order to emphasize local processing and reduce long anatomical connections. There is

already extensive evidence from fMRI and architectonic studies that the human brain contains many more cortical areas (possibly 150 or more) than the brains of macaque monkeys. One would also expect areas to be more frequently subdivided into sets of modules or columns of functionally related neurons, such as the three types of bands of neurons subdividing the second visual area, V2, of primates, but there is only limited evidence for this. However, the premise that hemispheric asymmetries increase with brain size in order to reduce the need for costly interhemispheric communication is well supported by the extensive evidence for hemispheric specializations in the human brain related to handedness, language, attention, memory and object recognition.

In conclusion, an outline of the course of the evolution of the human brain is starting to emerge. Great progress in the gathering of relevant data has occurred over the last 20–30 years. A more complete description could easily occupy a series of volumes, and interested readers are invited to read further. Most importantly, there is much yet to be gained by applying current methods of investigation so that future reviews can be better informed and greatly enriched.

Further reading

- Finlay, B.L., Darlington, R.B., and Nicastro, N. (2001). Developmental structure in brain evolution. *Behav. Brain Sci.* 24, 263–308.
- Kaas, J.H. (2000). Why is brain size so important: Design problems and solutions as neocortex gets bigger or smaller. *Brain and Mind* 1, 7–23.
- J.H. Kaas, ed. (2006). *Evolution of Nervous Systems*. 4 volumes. (London: Elsevier).
- Kaas, J.H. and Preuss, T.M. (2003). Human Brain Evolution. In *Fundamental Neuroscience*, Second Edition, L.R. Squire, ed. Elsevier, pp. 1147–1166.
- Kielan-Jaworowska, Z., Cifelli, R.L., and Luo, Z.X. (2004). *Mammals from the Age of Dinosaurs*. Columbia University Press.
- M.J. Ravosa and M. Dagosto, eds. (2006). *Primate Origins: Evolution and Adaptations*. Springer.
- Striedter, G.F. (2005). *Principles of Brain Evolution*. Sinauer.
- Wood, B. (2005). *Human Evolution: A Very Short Introduction*. Oxford University Press.

Department of Psychology, Vanderbilt University, 301 Wilson Hall, 111 21st Ave. South, Nashville, TN 37203, USA.
E-mail: jon.h.kaas@vanderbilt.edu

Correspondences

Chronic jet-lag increases mortality in aged mice

A.J. Davidson, M.T. Sellix, J. Daniel, S. Yamazaki, M. Menaker and G.D. Block

Despite the fact that trans-meridian travel and shift work are commonplace in our 24/7 society, few controlled studies have addressed the health effects of repeated phase shifts of the biological clock. Shift work [1] and chronic jet-lag [2] reduce mental acuity and increase the risk of a number of medical problems, including cancer, peptic ulcers and sleep disorders. Some of these problems become more severe with the number of years on the job, the result either of cumulative damage or the increased age of the subjects [3]. In general, morbidity associated with many organic disorders is increased in the aged; however, the role played by age-associated alterations in the circadian clock is poorly understood. In particular the effect of repeated schedule changes is largely unaddressed. Here we report evidence that chronic jet-lag increases mortality rates in aged mice.

We were led to the current experiment by an observation in an unrelated study where we found that three of eight aged transgenic rats exposed to a 6 hour advance of the light cycle died following the light schedule change. In contrast, no deaths were observed if the light cycle was delayed. In order to explore whether the effects of light schedule changes on longevity were reproducible in a larger study and observable in another rodent species, we placed young (8–12 month old) and aged (27–31 month old) C57BL/6 male mice on one of three lighting regimens

for eight weeks. Nine young and 30 aged mice were maintained on a normal 12:12 light–dark cycle. A second group of young ($n=9$) and old ($n=30$) mice was exposed to a 6 hour advance of the light-cycle once every seven days. The third group of young ($n=9$) and old ($n=28$) mice was phase-delayed by 6 hours once every 7 days. The rotating light schedules were chosen to effect large phase adjustments of the circadian system each week, such as would be expected to occur during flight across time zones or in some situations during rotating shift work cycles.

While younger mice fared well on this 8 week schedule (only one death occurred), we found that aged mice were significantly affected by light schedule changes (Figure 1A,B). At the end of the 8 week period of light schedule rotations there was 47% survival in animals whose light cycle was advanced each week, 68% in those experiencing delays of the light cycle and 83% in unshifted aged mice (chi square, all groups, $p<0.05$ on Day 54). Importantly, chronic stress was not implicated in this phenomenon as total daily fecal corticosterone levels did not increase in aged mice undergoing phase advances or phase delays (see Figure S1 in the Supplemental data published with this article online).

To determine whether the effects of phase advances on mortality might be related to the duration between schedule changes, mice were shifted more rapidly, every 4 days. On this schedule, we found that advanced mice died more quickly than with weekly shifts (Figure 1C; 60% survival on Day 24). Delays fared much better than advances (chi square $p<0.05$ on Day 32). The data suggest that the asymmetry in mortality rates between animals exposed to light schedule advances and delays persists and is possibly enhanced with the shorter inter-shift interval of 4 days.

Our data show that phase-advancing the light cycle hastens the death of aged mice. The mechanism underlying the

Gazetteer

L'Association pour la recherche sur le cancer

Otherwise known as ... l'ARC (pronounced 'lark').

What is it famous for? Being one of France's largest medical research charities and the best known, as far as the French public is concerned, because of its strong media presence.

Does it have much money to give away? L'ARC's income in 1999 was FFfr317 million (£32 million), three quarters of which went into funding research. Although this might not seem a lot, l'ARC — together with the other major French cancer charity, La Ligue Nationale Contre le Cancer (LNCC) — has an important role in plugging the gaps left by the stagnant budgets of CNRS and INSERM, France's public research agencies. Indeed, there are few biomedical scientists in France who have not benefited from l'ARC funding at one time or another.

So it's not all 'cancer research' then? L'ARC is considered very progressive and is known for its willingness to fund basic research that might seem far-removed from the cancer patient. To give an idea of l'ARC's breadth of coverage, this journal published 15 research articles that originated from French labs last year; 11 of these, with subject material ranging from neuron specification in frogs to blastomere fate in the zebrafish, acknowledged l'ARC for part-funding.

What else does l'ARC spend its money on? Last year, 410 stipends for graduate students and postdocs were disbursed by l'ARC to the tune of FFfr50.6 million (£5.2 million). These stipends, or *bourses*, are hard to come by in France and l'ARC's support in this area is seen as a lifeline,

particularly as French graduate students almost invariably take four years to complete their doctoral studies, and government support lasts only three years.

Where does the money come from? L'ARC raises almost all its money from public donations and legacies, relying heavily on its media presence. The charity is firmly associated in the public mind with the founder of l'ARC, Jacques Crozemarie (or Crozy as he is familiarly known), who until recently made regular televised appeals (in a white lab coat, although he has never been a physician or scientist). Advertisements in the press have been accompanied by signed photographs of Crozy.

How did l'ARC begin? L'ARC's forerunner, created in 1962 by Crozemarie, was the fund-raising body for the Institut Gustave-Roussy and several other cancer research labs based in the Paris suburb of Villejuif. It became a national charity in 1979 and since then has been the subject of some well-publicised controversies.

Why is it in the news? Crozemarie is alleged to have embezzled FFfr300 million (£30.6 million) of the charity's funds. He appeared in court in June — along with 25 others — accused of diverting the funds into Swiss bank accounts through an elaborate scheme of false and inflated invoices. The scandal, which was made public in December 1994, has been described by the French media as "the biggest racket involving a charity this century." In 1993, before its name fell into disrepute, l'ARC's income was FFfr581.2 million (£59.3 million). But according to France's national audit commission, which has been investigating the charity's accounts, less than one-third of l'ARC's income was spent funding research in 1993. Public generosity towards medical charities in general dropped sharply when the scandal broke.

Has l'ARC recovered? The installation of Michael Lucas — the former head of the watchdog body that exposed the financial irregularities of the charity — as the new president of l'ARC in 1996 and a purge of l'ARC's administrative council, seem to have helped rebuild public confidence, if the steady increase in donations is anything to go by.

What will happen to Crozemarie? He could face imprisonment for up to five years but his fate hangs in the balance as the French judiciary, like the rest of the country, grinds to a halt for the summer holidays.

Primer

Human memory Alan J. Parkin

At one level we can say that we know exactly what human memory is — the learning of new information involves changes in the synaptic connections between neurons specialised for the storage of information. This is known from *in vivo* studies of simple invertebrates and there is no reason to suppose that these fundamental mechanisms do not also apply to humans. We also know which regions of the brain are specialized for memory and that various neurotransmitters are involved in memory, notably acetylcholine, the neurotransmitter known to be massively depleted in Alzheimer's disease.

But how does this anatomical and physiological knowledge enable us to understand the different types of memory, or how any given piece of knowledge is derived from a particular combination of synaptic connections? Quite simply we have no idea at present and for this reason we have to think of memory at two levels: biological and psychological. Eventually, the two sources of

knowledge might merge into a single account of memory but at present the gap between the two areas of research is enormous.

Neuroanatomy of human memory

Damage to two main regions of the brain can cause severe loss of memory, or amnesia. The first of these is the hippocampal formation (see Figure 1). Operations in which its structures have been removed to cure epilepsy cause amnesia. More convincingly, patients who have suffered small strokes affecting only the hippocampal formation experience severe amnesia.

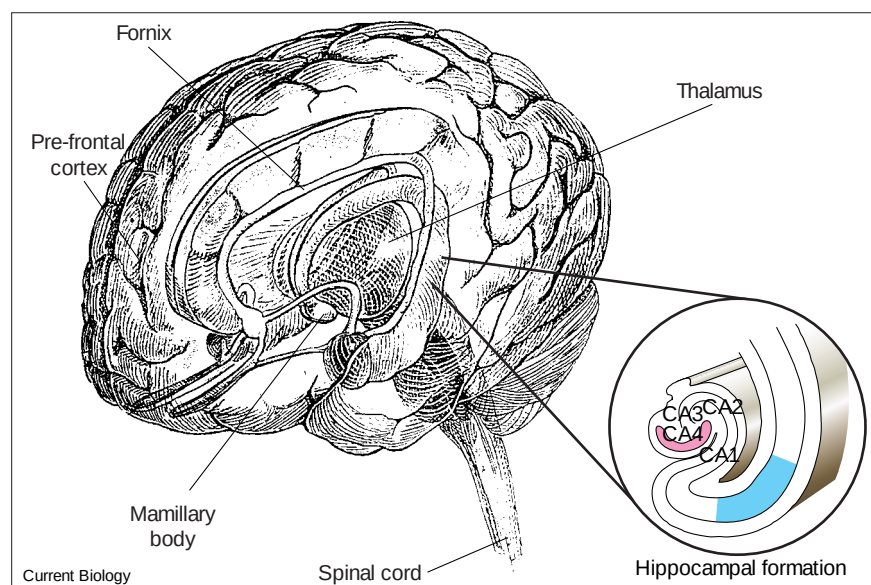
A second brain region concerned with memory is the diencephalon. Haemorrhaging in the diencephalon (in Korsakoff's Syndrome) results in dramatic loss of memory; strokes in this region can also have the same effect. The hippocampus has a large efferent pathway, known as the fornix, which terminates in the mamillary bodies. Lesions of the fornix have also caused significant loss of memory.

The various structures mentioned so far are all part of a circuit known as the limbic system (Figure 1). The fact that damage to any of the structures can cause amnesia has led many to propose that memory is mediated by a 'circuit' in which, rather like an electrical circuit, damage anywhere will have the same effect.

A psychological model of memory

Although there are some disagreements as to the psychological basis of memory, there is broad agreement on a 'modal' model for memory. The model is essentially based on information processing, where memory is conceived of as a series of 'stores' between which information flows. Three types of basic store are identified: sensory store, short-term store (STS), and long-term store (LTS) (see Figure 2). (These terms should not be confused with the commonly used phrases, 'short-term memory' and 'long-term memory', which memory theorists

Figure 1



The limbic system of the brain. The hippocampal formation comprises the hippocampus itself (which has four fields, CA1 to CA4) and associated structures including the dentate gyrus (pink) and the entorhinal cortex (blue). These structures all lie within the medial temporal lobe of the cerebral cortex. The other areas concerned with memory lie in an area called the diencephalon, these are the mamillary bodies, the dorso-medial nucleus of

the thalamus and the mamillothalamic tract. The fornix links the hippocampus and the mamillary bodies. The circuitry of the limbic system is completed by projections from the dorso-medial nucleus to the pre-frontal cortex, from which efferent pathways return to the hippocampal formation. (Brain diagram adapted with permission from Bloom FE, Lazerson A: *Brain, Mind and Behavior*, 2nd edn. New York: WH Freeman; 1988.)

avoid because they are used very imprecisely to indicate periods of time over which information is held. In contrast, STS and LTS refer to hypothetical brain structures with assumed properties. 'Working memory' is another term that often crops up — for the purpose of this article working memory is best considered to be a more detailed aspect of STS.)

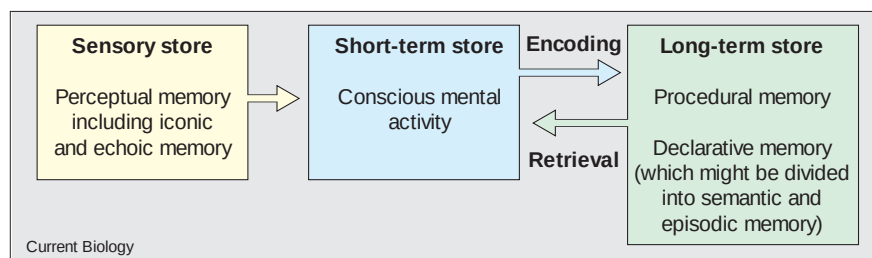
The sensory store provides short-lived storage of the perceptual input — visual information, known as 'iconic memory', is thought to last about half a second, whereas auditory information, 'echoic memory', lasts several seconds. Information then passes into STS; this represents the locus of conscious mental activity in which incoming information interacts with previously stored knowledge to direct a response.

Previously stored knowledge resides in LTS. Information enters

via a process known as 'encoding' and, once in LTS, is accessed by a process called 'retrieval'. STS has a very limited capacity — this is indicated by the digit-span test in which subjects are required to repeat back a series of random digits immediately. Adults typically manage about seven digits before they start making mistakes. In addition the information in STS is actively stored, rather like the memory of a calculator, and can be lost if there is some form of distraction. Information in LTS is thought to be passively stored because it has undergone the process of 'consolidation' in which new memories come to be represented in terms of new patterns of synaptic interconnections.

Biological evidence

Patients with amnesia provide good evidence for the distinction between

Figure 2

The current modal model for memory, in which information flows between a series of stores.

STS and LTS. Their performance in the digit-span test, which requires only STS, is normal but performance in the story recall test, with its demands on LTS is very poor. Similar effects can be demonstrated in normal people given the drug scopolamine, which impairs action of acetylcholine; patients given scopolamine perform normally on digit-span but have markedly impaired story recall.

Biological evidence also bears on the distinction between encoding and retrieval. Amnesia takes two basic forms: 'anterograde', in which there is an inability to remember new information, and 'retrograde', in which the patient fails to remember information acquired before their brain injury. If encoding and retrieval were linked in some essential way we would expect these two forms of amnesia to be correlated with one another but they are not, suggesting independent mechanisms. Data from normal humans also support this. The drug imipramine when given to normal volunteers impairs the learning of new information but does not interfere with retrieving information learned before the drug was administered.

The nature of LTS

LTS has been described as if it were some single, large structure containing all our stored knowledge. In fact, there is abundant evidence that LTS has different independent components. There is a proposed division between so-called 'procedural' and 'declarative'

memory. Procedural memory is defined as information we possess but which we cannot describe verbally. Procedural memory is typified by skills such as typing, and playing a musical instrument. Thus when a typist is asked how to type the word 'caterpillar' he or she will make the finger movements to provide the answer.

It is important to note that procedural memory is not a single entity; rather, a number of memory skills have a procedural quality. We know this because comparisons of people's ability on different forms of procedural memory task do not correlate — which they should if procedural memory were a single structure.

Declarative memory represents all knowledge that can be consciously accessed and expressed symbolically through speech and writing. There is a lively debate as to whether declarative memory is a single system or whether it has two separate components: 'semantic memory' — our knowledge of language, concepts or facts — and 'episodic memory' — our personal autobiographical knowledge of the past. Two lines of evidence bear on this issue: evidence from the relatively new field of functional neuroimaging (see green box) and the breakdown of LTS in human amnesia.

Amnesia and LTS

Studies of amnesic patients have shown that they perform well on various tests of procedural memory. Thus, as well as showing the

preservation of existing skills, amnesic patients have also been shown to acquire new skills (for example, typing). The perceptual learning of amnesic patients — the ability to analyse and identify new visual forms — also seems to be normal. As amnesia is typically caused by damage to the limbic and frontal lobe structures, this evidence strongly indicates that memory systems with procedural characteristics are not mediated by structures in the limbic system.

By contrast, declarative memory is impaired in amnesic patients; they have immense problems retaining any form of factual or event-based knowledge. The limbic and frontal structures must therefore be directly concerned with declarative memory.

But what about the finer distinction between episodic and semantic declarative memory? It is well known that amnesic patients fail to acquire new vocabulary (the much studied amnesic patient HM has learned only six new words since his temporal operation in 1953) as well as failing to remember ongoing personal events. This seems to support the idea of a single declarative memory system. But it might be that episodic memory helps us to acquire new semantic memory. For example, your recall of the word 'hippocampus' might depend on your remembering reading this article.

Crucial evidence in this debate has centred on the nature of retrograde amnesia. Various researchers have claimed that retrograde amnesia is principally a failure of episodic memory. Although it is true that memory for episodes is severely impaired in amnesia there is now good evidence that the semantic type of memory is also affected. Thus, retrograde amnesia results in a failure to recognise people from the past and failure to define words and concepts introduced before the brain injury. Interestingly both these kinds of impairment become more severe for information acquired closer to the brain injury — a phenomenon

known as the 'temporal gradient'. Explanation of this phenomenon remains a mystery although some have suggested that it reflects the time-course of consolidation. The fact that temporal gradients can extend back 20 years makes this a difficult explanation to accept.

The roles of different structures

Limbic and frontal structures are best thought of as mediating declarative memory but what role do these different structures play in the process of memory? The hippocampus is divided into four 'fields' and one of these, CA1, receives inputs from all parts of the prefrontal cortex. Moreover, neurons in CA1 have a very high degree of interconnectivity. Memory for an

event is thought to represent the integration of different aspects of information and, although there is no direct proof, it is thought that CA1 provides the anatomical basis and that other parts of the hippocampal formation 'fix' a particular event and then mediate storage of that event in the temporal cortex.

Far less is known about the function of diencephalic structures in memory but one idea is that these are involved in encoding information which enables the time-based characteristics of a memory to be established. The frontal lobes are generally thought to be involved in more 'strategic' aspects of memory. These strategic processes involve the operations used to determine what aspects of an event are encoded and those that control how memories are retrieved. Thus, studies of patients with frontal lesions have shown that their free recall of information is often poor but improves dramatically with prompting. Also their learning processes often reveal a failure to attend to the most relevant aspects of an event. There is also evidence that the pre-frontal cortex is crucial for determining the truth of memories because, quite often in frontal lobe damage, patients suffer from 'illusory memory' or 'confabulation' in which they tend to remember things that have not actually happened.

It should be obvious that the study of human memory is complex, not least because the problems can be approached on two fundamentally different levels — the biological and the psychological. There is progress in both these domains but it will be a long time before they come together to provide a single explanation of human memory.

Key references

- Baddeley AD: *The Psychology of Human Memory*. Hove: Psychology Press; 1997.
 Baddeley AD: *Essentials of Human Memory*. Hove: Psychology Press; 1999.
 Parkin AJ: **The hippocampus is the key.** *Curr Biol* 1996, **6**:1583-1585.
 Parkin AJ: *Memory and Amnesia: an Introduction*. Hove: Psychology Press; 1997.

Parkin AJ: **Novelty, association and the brain** *Curr Biol* 1997, **7**:R768-R769.
 Squire LR: *Memory and Brain*. New York: Oxford University Press; 1987.

Address: Experimental Psychology, University of Sussex, Brighton BN1 9QG, UK.

Correspondence

A latrophilin/ CL-1-like GPS domain in polycystin-1 C.P. Ponting*, K. Hofmann† and P. Bork‡,§

Autosomal dominant polycystic kidney disease (ADPKD) is a common inherited disorder (incidence of 1 in 1,000), the cardinal manifestations of which are renal and liver cysts and intracranial aneurysm. The gene defective in the most common and severest form of ADPKD, *PKD1*, is broadly expressed and encodes a 4,302 amino acid plasma-membrane protein, polycystin-1 (PKD1) [1–3]. Despite detailed knowledge of the domain architecture for the majority of the PKD1 sequence (Figure 1), little is known of PKD1 function.

Identification of *PKD1* mutations has been hampered by the presence of numerous *PKD1* homologues elsewhere on chromosome 16 [2] and *in vitro* expression of full-length PKD1 is yet to be reported.

Here, we report the identification of three previously unrecognised domains in PKD1 that are likely to possess distinct carbohydrate-, lipid- and protein-binding functions. These domains were identified using PSI-BLAST database searches [4,5] with an expect-value (*E*-value) inclusion threshold of *E* < 0.01. Independent evidence was provided using a generalised profile analysis method [6], in which the significances of findings were better than *p* < 0.01 in

Functional neuroimaging

In functional neuroimaging, a range of different techniques are used to indicate which parts of the brain are receiving most blood flow and are, hence, most active. Normal people are given memory tasks while undergoing functional neuroimaging. Such studies have confirmed much of what was already known about the neuroanatomical basis of memory but the evidence has also thrown up some new issues and addressed some controversies. First, neuroimaging studies have indicated different patterns of activation when people are answering questions about episodic and semantic memory — contradicting the evidence from amnesia. Studies have also highlighted the role of the frontal lobes of the brain in encoding and retrieval, although the original view that the left frontal region mediated encoding and the right mediated retrieval is now in some doubt. A more robust finding is the demonstration of frontal lobe involvement in the setting up of new links between objects (known as 'association') and the hippocampal involvement in detecting that an object is novel ('stimulus novelty'). Most interesting, perhaps, has been the discovery that a little-known parietal lobe structure, the right precuneus, is active during retrieval — a finding that is stimulating various ideas about how the various elements of a memory might be brought together during the process of retrieval.

Primer

Information theory in the brain

Pamela Reinagel

Claude Shannon's classic 1948 paper introduced a general theory for measuring the transmission of information from a source, across a noisy channel, to a receiver (Figure 1). This theory became known as *information theory*. Shannon illustrated his theory with such examples as messages sent over telegraph lines by Morse code. But the same principles can be applied to neurons: a neuron transmits information along its axon to other neurons, using a neural code. Information theory has proved to be a powerful tool for quantifying the communication of information by neurons.

Bits

Shannon argued that information should be measured in the now-familiar units of bits. A bit is a binary digit (0 or 1), or the amount of information needed to give an answer to a single yes-or-no question. The outcome of one coin toss could be reported in one bit (0 for heads, 1 for tails). The result of two coin tosses could be reported in two bits, which can represent four possible combinations (00, 01, 10, 11). In general, with n coin tosses there are $N=2^n$ possible outcomes. Turning this around, if there are 2^n possible combinations or messages, then there are $\log_2(2^n) = n$ bits of information. The reason $\log_2$ shows up so often in information theory is that the unit of information is the binary (base 2) digit.

Information content

The information content in a source can be measured as the minimum number of bits that could represent the messages from that source. This

quantity is called the *entropy* of the source, because its equation has the same form as the entropy in statistical thermodynamics. In the simplest case, when there are N possible messages and all are equally likely, the entropy is $\log_2 N$, as for the coin tosses above.

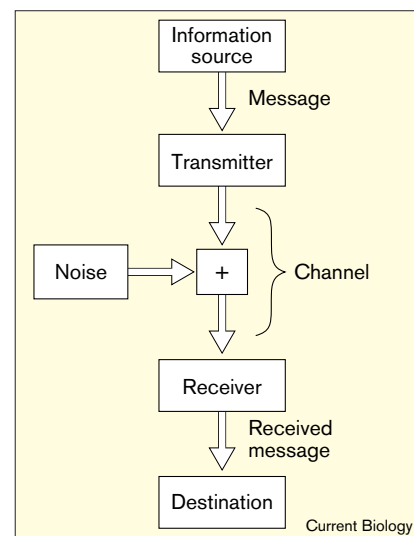
The situation is more complicated if two coins are weighted, so that each of them comes up heads in 95% of tosses. At first, it seems we still need two bits to represent the four possibilities. But we could instead use a short code, 1, to designate the usual outcome, and longer codes for the three uncommon outcomes, for example 0 as a prefix for rare outcomes followed by 00, 01 or 10 to distinguish them. This scheme would use 1 bit $0.95 \times 0.95 \approx 90\%$ of the time, and 3 bits 10% of the time, hence on average using 1.2 bits instead of 2. But is 1.2 the minimum number of bits that could represent this source?

According to information theory, a message (outcome) that occurs with probability P has an entropy of $\log_2(1/P)$ bits. For a simple coin toss, each outcome has a probability $P=1/2$ and thus an entropy of $\log_2(2) = 1$ bit. In the case of the two weighted coins, the outcome two heads is so likely that its actual occurrence does not contain much new information: $P=0.9$, and $\log_2(1/0.9) \approx 0.15$ bits. By contrast, the outcome two tails would be very surprising, and thus highly informative: $P=1/400$, and $\log_2(400) \approx 8.6$ bits. Shannon proved that the entropy of an information source is the sum over all possible messages of the entropy of the message, $\log_2(1/P)$, weighted by how often it occurs (P):

$$\text{Entropy} = \sum P \log_2 \left(\frac{1}{P} \right)$$

In the case of the two weighted coins above, the entropy turns out to be only 0.6 bits. This means an optimal code could be twice as efficient as the code proposed above, which used

Figure 1



Shannon's schematic of a general communication system. An information source produces a message, which is then transmitted over a noisy channel to the receiver. The message received at the destination is a corrupted version of the original message.

1.2 bits. Note that the entropy of the two weighted coins is much less than two fair coins (2 bits). In general, a system has maximum entropy when every possible message is equally likely. When probabilities are unequal, the entropy is always reduced. At the extreme, when one message has a probability $P=1$ and all others $P=0$, the entropy is 0.

Finally, we have been assuming that each message is independent of the others. Entropy is reduced if the probability of a message depends on previous ones. For example, the letter u is fairly uncommon in English text, but after a q it is almost always present. Thus the letter u has high information in general, but almost no information when following q . When messages are not independent, it is necessary to measure the probabilities of sequences of messages to determine the entropy of the source.

Information transmission

Between the source and the receiver, information may become

corrupted by noise (Figure 1). The amount of information that gets through can be measured by how much information the sent and received messages have in common — their *mutual information*. To choose an example from neuroscience, a visual display on which a stimulus is shown could be considered an information source. All the optical and neural events leading up to the firing of a neuron could be taken as the channel, and the neuron's firing rate could be taken as the message received at the destination. We can then ask: how much information about the stimulus is represented by the cell's firing rate?

Suppose the stimulus, S , has two possible values — black and white —chosen randomly on each experimental trial with equal probability. The minimum number of bits required to represent this stimulus is 1 bit, so this is the stimulus entropy. One could classify the cell's response, R , as either on or off in each trial, and choose the threshold in such a way that the cell is considered on in exactly half the trials. Defined this way, the response also has an entropy of 1 bit.

To find out how much information this response contains about the stimulus, we tabulate the probability of each combination of S and R (Figure 2). The response of the neuron in Figure 2a is completely independent of the stimulus, as we might find if we were recording from a purely auditory neuron in response to this visual stimulus. The probability of any combination of stimulus and response is the same as expected by chance. Thus, even though the response has 1 bit of entropy, it contains no information about the stimulus. The response of the neuron in Figure 2b, however, is a perfectly reliable indicator of the stimulus. It contains all the information in the stimulus, 1 bit. A noisy visual neuron, whose

Mutual information

Shannon showed that the mutual information between R and S can be calculated from the joint probability distribution (Figure 2) by the equation:

$$I = \sum_S \sum_R P(S,R) \log_2 \left(\frac{P(S,R)}{P(S) \times P(R)} \right)$$

This seems more complicated than it is. The joint probability of stimulus S and response R (one shaded square) is written $P(S,R)$. The ratio compares this joint probability to what might happen by chance — the product of the two individual

probabilities $P(S) \times P(R)$. If these two are equal, as for the neuron in Figure 2a, then the ratio is 1, and $\log_2(1) = 0$ bits. If the joint bits probability differs from chance, information is encoded. For example, for the neuron in Figure 2b the joint probability of getting an on response with a black stimulus is 0.5, compared to the expected $0.5 \times 0.5 = 0.25$, so the ratio is 2, and $\log_2(2) = 1$ bit. Finally, $\sum_S \sum_R P(S,R)$ simply indicates taking a sum over all stimuli and all responses, weighted according to how often the combination occurs.

Figure 2

(a)

		Response	
		Off	On
Stimulus	White	$p=0.5$	$p=0.5$
	Black	$p=0.5$	$p=0.5$

(b)

		Response	
		Off	On
Stimulus	White	$p=0.5$	$p=0.5$
	Black	$p=0.5$	$p=0.5$

Current Biology

Responses of two neurons during presentation of a random, binary stimulus. The neuron in (a) codes no information, whereas the cell in (b) is a perfect encoder.

response was partially determined by the stimulus, would fall between these extremes.

In most neurobiology experiments, both the stimuli and the responses are more complicated than this example. Some stimuli have hundreds or thousands of bits of information per second. For example, a stimulus could have many distinct shades of gray, with different shades at different spatial locations, and the shades could be changed hundreds of times per second, to make a black-and-white movie. Neural responses are also much richer than simply on and off. For example, a response could be defined by the exact time of each action potential fired by each cell in a large population of sensory neurons. Even for fairly simple

experiments, it is usually impossible to collect enough experimental data to fill in a table like those in Figure 2. Fortunately, several sophisticated tricks have been devised to estimate the mutual information in other ways (for one example, see Figure 3).

Cracking neural codes

Information theory has been used to study how neurons encode information. Firing rates have long been known to be important for neural codes; the firing rates of single neurons often correlate systematically with experimental parameters we can vary or measure in the laboratory, such as the intensity of a sensory stimulus or the magnitude of a muscular contraction. Using information theory, we can put

a number on the amount of information encoded by the firing rate of a neuron. We can then ask whether additional information is transmitted by other features of the neural response.

The most striking lesson to emerge over the past decade is that the exact timing of action potentials is important in neural coding, particularly when the encoded signal is itself rapidly varying. For the most part, this is an extension of the idea that firing rates code information, but we now think of a neuron's rate (probability of firing) as varying on the timescale of milliseconds rather than seconds. There has been intense interest in whether information is encoded in other aspects of neural responses, such as temporal patterns of firing, or firing patterns involving multiple cells. In recent years, a few examples of such temporal and population coding

have been demonstrated, although so far the amount of information involved has been modest. It remains controversial whether these codes will turn out to be functionally important.

From information theory, we know that the most efficient code to transmit a signal depends on the statistics of the signal. This is why Morse code uses a short symbol (•) for the frequent letter *e*, and a clumsy long one (— • —) for the rare *q*. But for a language with few *es* and many *qs*, this would be an especially bad code. Similarly, a neural code that represented one class of sensory stimuli efficiently would necessarily be inefficient for other stimuli. It has been proposed that sensory neurons evolved to send as much information to the brain as possible under real-world conditions. Early tests of this theory have confirmed the prediction that sensory neural codes seem to be

specifically adapted for the statistics of their natural stimuli.

Broader applications

Information theory is a completely general method to measure the transfer of information from one place to another, and is particularly suited to describing many aspects of neural function. Information theory has been applied most widely in the field of sensory coding, but it is equally applicable to the transmission of neural commands for motor output patterns. One can also measure the mutual information between the spike train of one neuron and that of its postsynaptic target, or between a single cell's synaptic currents and its action potentials.

The term *information* suggests to us something about the intention of a sender or the value to the receiver. But, in the uses of information theory presented above, neither is implied. For example, there is more information content (entropy) in a randomly flickering visual stimulus than in a real-world scene, even though the flickering is meaningless. In the future it will be important for the field to address such additional issues as: the value of different stimulus information for the animal; the role of active exploration of the environment in selection stimuli for encoding; and the mechanisms in the brain for decoding the information carried by neural responses.

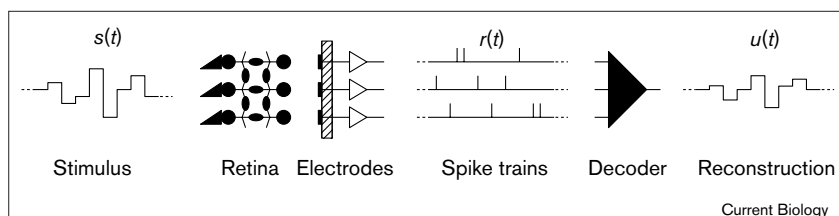
Decoding a neural code

One approach to studying neural codes is to try to decode a neural response and reconstruct the original signal.

Any information about the stimulus that can be successfully reconstructed is thereby demonstrated to be encoded by the neuron(s). The successful decoding algorithm may also provide insight into how that information is represented, and how other neurons might extract it. For example, suppose one has recorded the spike trains of several visual neurons in

response to a randomly modulating stimulus (Figure 3). One could try to find a decoder that would estimate the visual stimulus from the responses of the cells. With judicious choice of the input signal, efficient methods can be used to compare the reconstruction to the original signal, to obtain an information estimate without recourse to a joint probability table (Figure 2). In this way mutual information can be estimated with comparatively little data.

Figure 3



The luminance of a visual stimulus $s(t)$ is chosen independently at each time step. The responses of retinal ganglion cells $r(t)$ are recorded on multiple electrodes. A decoder is sought that can produce a reconstruction, $u(t)$, that matches the original stimulus as well as possible. (Adapted with permission from Warland DK, Reinagel P, Meister M: *J Neurophysiol* 1997, **78**:2336-2350.)

Key references

- Shannon CE: **A mathematical theory of communication.** *Bell System Tech J* 1948, **27**:379-423;623-656. (Reprinted in *Claude Elwood Shannon: Collected Papers*. Edited by Sloane NJA and Wyner AD. New York: IEEE Press; 1993. Also available at <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>)
- Rieke FM, Warland DK, de Ruyter van Steveninck R, Bialek W: *Spikes: Exploring the neural code*. Cambridge, Massachusetts: MIT Press; 1997
- Cover TM, Thomas JA: *Elements of Information Theory*. New York: Wiley; 1991.

Address: Department of Neuroscience, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

Primer

Motor prediction

Daniel M. Wolpert* and J. Randall Flanagan†

The concept of motor prediction was first considered by Helmholtz when trying to understand how we localise visual objects. To calculate the location of an object relative to the head, the central nervous system (CNS) must take account of both the retinal location of the object and also the gaze position of the eye within the orbit. Helmholtz's ingenious suggestion was that the brain, rather than sensing the gaze position of the eye, predicted the gaze position based on a copy of the motor command acting on the eye muscles, termed efference copy. He used a simple experiment on himself to demonstrate this. When the eye is moved without using the eye muscles (cover one eye and gently press with your finger on your open eye through the eyelid), the retinal locations of visual objects change, but the predicted eye position is not updated, leading to the false perception that the world is moving. The concept of efference copy was firmly established by experimental work of Von Holst and Sperry in the 1950s. Since then the idea that we predict the consequences of our motor commands has emerged as an important theoretical concept in all aspects of sensorimotor control.

In general, prediction refers to estimating future states of a system. The extent to which our CNS can influence these future states covers a continuous range. For example, most people have little influence on the behaviour of the stock market, but a great deal of influence over the behaviour of their own body. We consider motor prediction as representing predictions of systems which are directly and usually

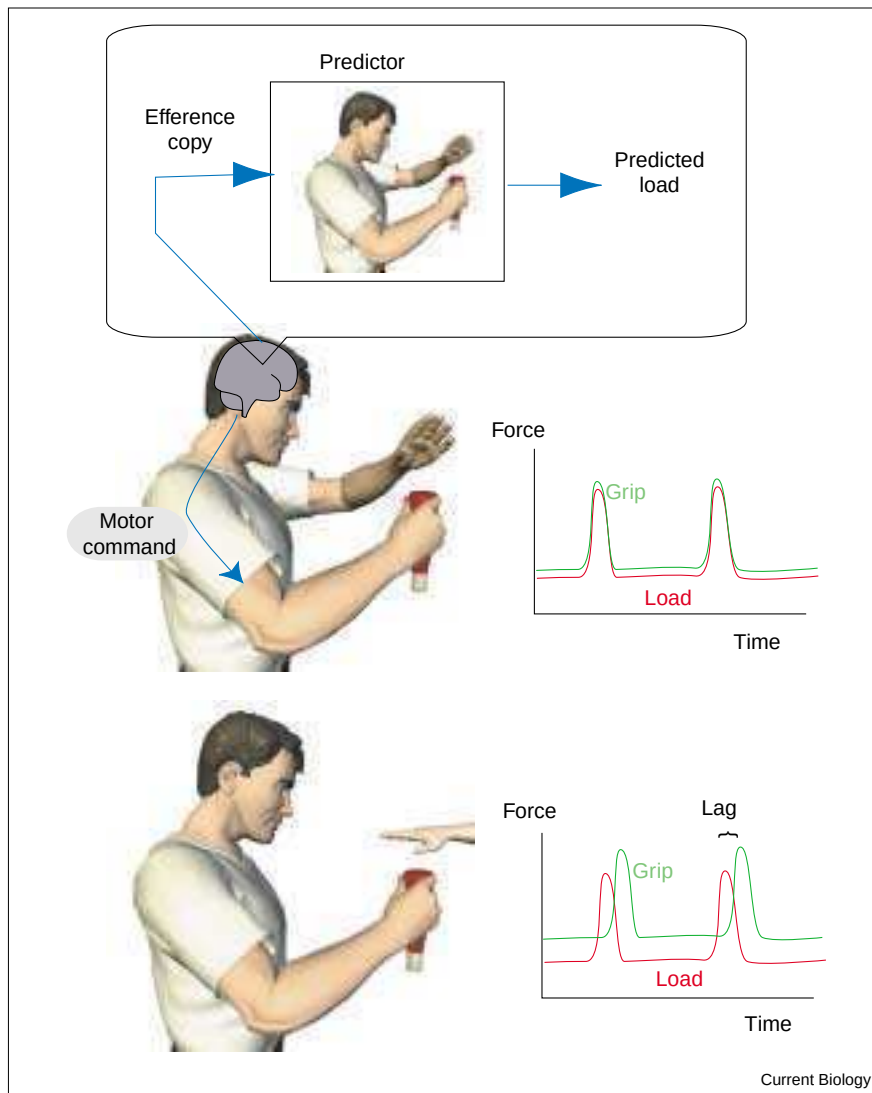
immediately influenced by our motor commands. Therefore, motor prediction could predict how our arm moves in response to a motor command, as well as how a car we are driving responds to our foot movement on the pedals. The relationship between a motor command sent to the eye and the consequences for eye position are relatively simple. However, when we consider limb motion the relationship between our motor commands and behaviour is far more complex due to the dynamics of multi-joint motion. This complexity is even greater when we consider the plethora of interactions that our bodies can engage in with the environment, such as tool use. To predict the consequences of a motor command in such complex situations requires a system that can simulate the dynamic behaviour of our body and environment. Such a system is termed an internal forward model as it is internal to the CNS, models the behaviour of the body and captures the forward or causal relationship between actions and their consequences.

Skilled motor behaviour relies on accurate predictive models of both our own body and tools we interact with. As the dynamics of our body change during development and as we experience tools which have their own intrinsic dynamics, we need to acquire new models and update existing models. Thus forward models are not fixed entities but must be learned and updated through experience. Forward models can be trained and updated using prediction errors, that is by comparing the predicted and actual outcome of a motor command. Well established computational learning rules can be used to translate these errors in prediction into changes in synaptic weights which will improve future predictions of the forward model. Here we review the uses of motor prediction in sensorimotor control.

State estimation

Knowing our body's state, for example the positions and velocities of our body segments, is fundamental for accurate motor control. However, sensory signals that convey information about state are subject to significant delays due to receptor transduction, neural conduction and central processing. Moreover, sensors are not perfect and provide information which is corrupted by random processes, known as noise. Using sensory information to estimate the state can lead to large errors especially for fast movements, and when this erroneous state is used to control movement it can lead to instability. An alternative is to estimate state using prediction based on motor commands. Here the estimate is made ahead of the movement and therefore is better in terms of time delays, but the estimate will drift over time if the forward model is not perfectly accurate. The drawbacks of both these mechanisms can be ameliorated by combining sensory feedback and motor prediction to estimate the current state. Such an approach is used in engineering and the system which produces the estimate is known as an observer, an example of which is the Kalman filter. The major objectives of the observer are to compensate for the delays in the sensorimotor system and to reduce the uncertainty in the state estimate which arises through noise inherent in both the sensory and motor signals. Such a model has been supported by empirical studies examining estimation of hand position, posture and head orientation.

Skilled motor behaviour involves different modes of control which rely on prediction and sensory feedback to different extents. These models of control are nicely illustrated within the context of object manipulation. When holding an object in a precision grip with the fingertips on either side, sufficient

Figure 1

To prevent a ketchup bottle from slipping, sufficient grip force must be exerted to counteract the load. When the load is increased in a self-generated manner (left hand strikes the ketchup bottle, top), a predictor can use an efference copy of the motor command to anticipate the upcoming load force and thereby generate grip which

parallels load force with no delay. However, when the load is externally generated (another person strikes the bottle, bottom), then it cannot be accurately predicted. As a consequence, the grip force lags behind the load force and the baseline grip force is increased to compensate and prevent slippage.

grip force must be generated to prevent slip due to load force exerted by the object. When the object's behaviour is unpredictable, sensory feedback provides the most useful signal for estimating load. For example, when flying a kite or holding the hand of a rambunctious child, we need to adjust our grip in

response to the unpredictable actions of the kite or child. When dealing with such unpredictable objects our grip force is modified reactively in response to sensory feedback from the fingertips, with the consequence that grip tends to lag behind load. However, when we direct behaviour towards objects in

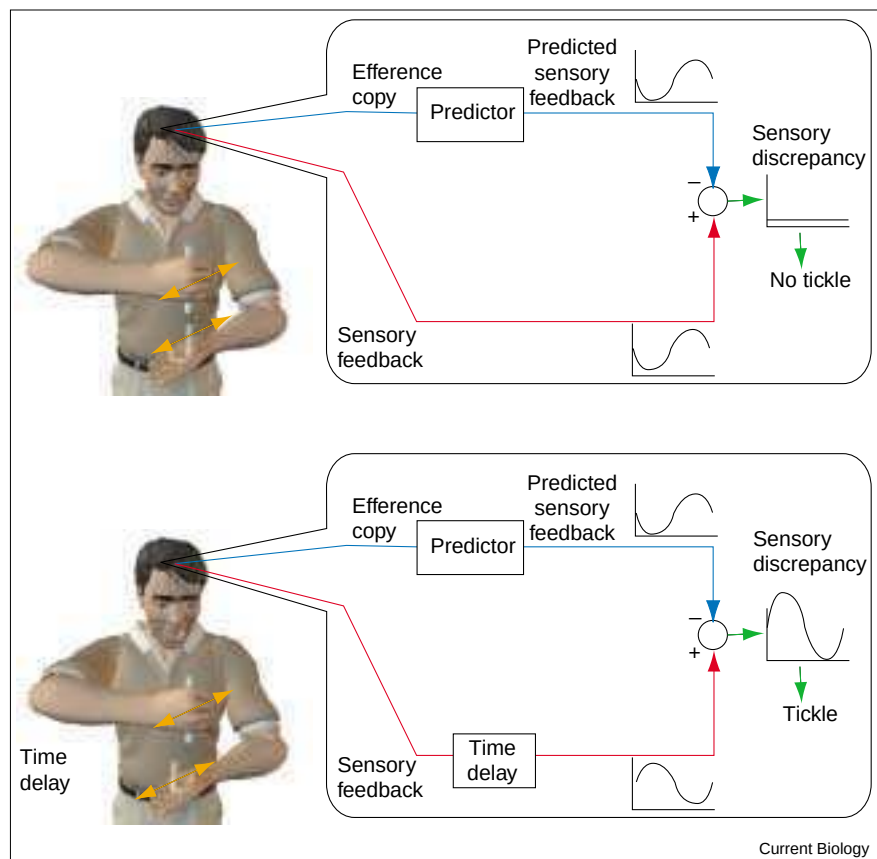
the environment that exhibit stable properties, predictive control mechanisms can be effectively exploited (Figure 1). For example, when the load is increased by a self generated action, such as moving the arm, the grip force increases in parallel with load force with no delay. Sensory detection of the load is too slow to account for this increased grip force which relies on predictive processes. Such predictive control is essential for the rapid movements commonly observed in dexterous behaviour.

Sensory confirmation and cancellation

In addition to state estimation, prediction allows us to filter sensory information, attenuating unwanted information or highlighting information critical for control. Sensory prediction can be derived from the state prediction and used to cancel the sensory effects of movement (reafference). By using such prediction, it is possible to cancel out the effects of sensory changes induced by self-motion, thereby enhancing more relevant sensory information. For example, predictive mechanisms underlie the observation that the same tactile stimulus, such as a tickle, is felt less intensely when it is self-applied. This mechanism has been supported by studies in which a time delay is introduced between the motor command and the resulting tickle (Figure 2). The greater the time delay the more ticklish the percept, presumably due to a reduction in the ability to cancel the sensory feedback based on the motor command.

Similarly, sensory predictions provide a mechanism to determine whether motion of our bodies has been generated by us or by an external agent. For example, when I move my arm, my predicted sensory feedback and the actual feedback match and I therefore attribute the motion as being generated by me.

Figure 2



An experiment in which subjects tickle themselves through a robotic interface (not shown). In the top Figure the motion of the right hand is directly transmitted to the left. Using a predictor the sensory consequences of this motion are estimated and subtracted from the actual sensory feedback leaving little discrepancy. In the lower Figure a time delay was introduced between the motion of the

right hand and the effect on the left. The motion of the right hand and stimulus on the left hand is the same as in the upper figure. However, the temporal rearrangement between cause and effect means that the prediction is out of synchronisation with the actual feedback, leading to a large sensory discrepancy which is felt as tickle.

However, if someone else moves my arm, my sensory predictions are discordant with the actual feedback and I attribute the movement as not being generated by me. Therefore, in general, movements predicted based on my motor command are labelled as self-generated and those that are unpredictable are labelled as not produced by me. Frith has proposed that a failure in this mechanism may underly delusions of control in schizophrenia, in which it appears to the patient that their body is being moved by forces other than their own. Interestingly, Sirigu and

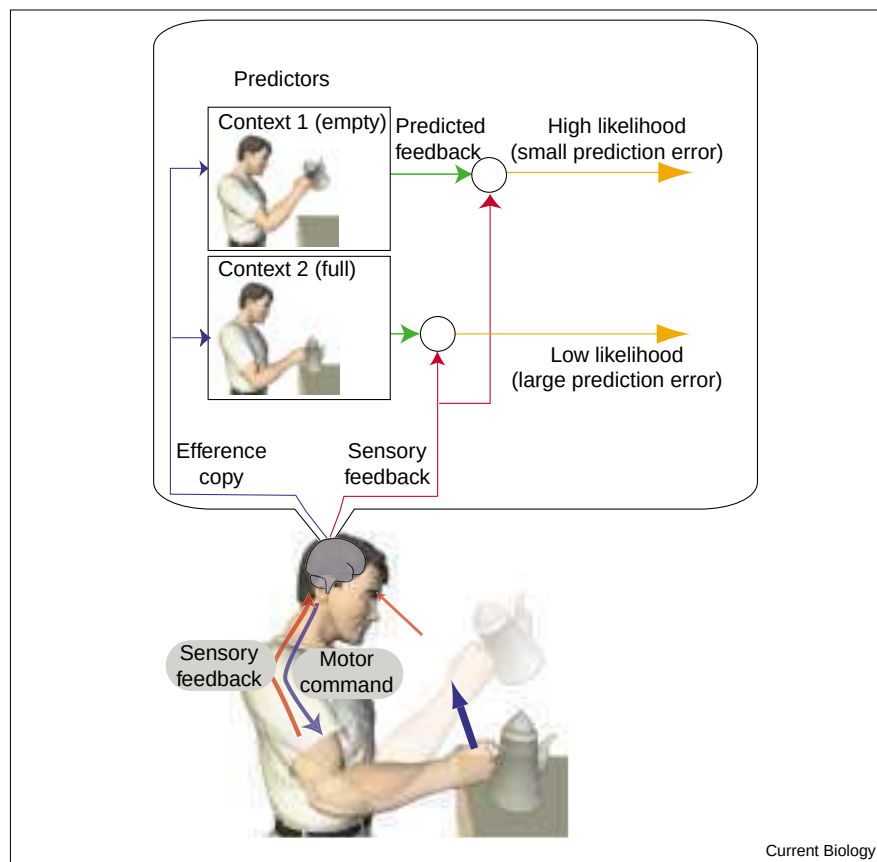
colleagues have shown that damage to the left parietal cortex can lead to a relative inability to determine whether viewed movements are ones own or not.

The discrepancy between actual and predicted sensory feedback is essential in motor control. For example, when we pick up an object we anticipate the timing of discrete events such as object lift off. The CNS is particularly sensitive to the occurrence of unexpected events or the absence of an expected event. Thus if the object is lighter or heavier than expected, and therefore

lifts off too early or fails to lift off, reactive responses are evoked. The CNS seems to specifically monitor these critical moments to confirm the progress of the task and to engage subsequent phases in the cascade of actions underlying natural tasks.

Context estimation

Humans demonstrate a remarkable ability to generate accurate and appropriate motor behaviour under many different and often uncertain environmental conditions. It has been proposed that the CNS uses a modular approach in which multiple controllers co-exist and are selected based on the movement context. Therefore when we pick up an object with unknown dynamics we need to identify the context and select the appropriate controller. One possible solution to this identification and selection problem has been proposed in the form of the MODular Selection and Identification for Control (MOSAIC) model. The idea is that, when lifting an object, the brain simultaneously runs multiple forward models that predict the behaviour of the motor systems when interacting with different previously learned objects (Figure 3). Each forward model generates a prediction of the sensory feedback that should be obtained for its context. Moreover, each forward model is paired with a corresponding controller forming a predictor-controller pair. If the prediction of one of the forward models closely matches the actual sensory feedback, then its paired controller will be selected and used to determine subsequent motor commands. In computational terms, the sensory prediction error from a given forward model is represented as a probability; if the error is small then the probability that the forward model is appropriate is high. The set of probabilities from an array of forward models is used to weight the

Figure 3

A schematic of context estimation with just two contexts, that a teapot is empty or full. When a motor command is generated, an efference copy of the motor command is used to simulate the sensory consequences under the two possible contexts. The predictions based on an empty teapot suggest that lift-off will take place early

compared to the full teapot context and that the lift will be higher. These predictions are compared with actual feedback. As the teapot is in fact empty, the sensory feedback matches the predictions of the empty teapot context. This leads to a high likelihood for the empty teapot and a low likelihood for the full teapot.

outputs of the corresponding controllers.

Mental practice, imitation and social cognition

Not only is prediction essential for motor control, it may also be fundamental for high level cognitive functions including action observation and understanding, mental practice, imitation and social cognition. The forward model may provide a general framework for prediction in all of these domains. In these situations a forward model is used to predict the sensory outcome

of an action, without actually performing the action. For example, it is conceivable that mental practice can improve performance by tuning controllers or selecting between possible mentally rehearsed actions. Imaging studies have shown that brain areas active during mental rehearsal of an action are strikingly similar to those used in performing the action. Moreover, the durations of mentally simulated movements are tightly correlated with the durations of actual movements, a correlation that is lost with damage to parietal cortex.

In motor control, a forward model can be used to predict the sensory consequences of our actions. In perception of action we could use multiple forward models to make multiple predictions and, based on the correspondence between these predictions and the observed behaviour, we could infer which of our controllers would be used to generate the observed action. Finally, in social interaction, a forward social model could be used to predict the reactions of others to our actions. It may be that the same computational mechanisms which developed for sensorimotor prediction have adapted for other cognitive functions.

Further reading

- Blakemore SJ, Frith CD, Wolpert DM: **Perceptual modulation of self-produced stimuli: the role of spatio-temporal prediction.** *J Cognit Neurosci* 1999, **11**:551-559.
- Flanagan JR, Wing AM: **The role of internal models in motor planning and control: evidence from grip force adjustments during movements of hand-held loads.** *J Neurosci* 1997, **17**:1519-1528.
- Frith CD, Blakemore SJ, Wolpert DM: **Abnormalities in the awareness and control of action.** *Phil Trans R Soc Ser B* 2000, **355**:1771-1788.
- Johansson RS, Edin BB: **Predictive feedforward sensory control during grasping and manipulation in man.** *Biomed Res* 1993, **14**:95-106.
- Sirigu A, Duhamel JR, Cohen L, Pillon B, Dubois B, Agid Y: **The mental representation of hand movements after parietal cortex damage.** *Science* 1996, **273**:1564-1568.
- Wolpert DM, Ghahramani Z: **Computational principles of movement neuroscience.** *Nat Neurosci* 2000, **3**:1212-1217.
- Wolpert DM, Ghahramani Z, Flanagan JR: **Perspectives and problems in motor learning.** *Trends Cognit Sci*, in press.

Acknowledgements

This work was supported by grants from the Human Frontier Science Project, the Wellcome Trust, the Natural Sciences and Engineering Research Council of Canada and the BBSRC.

Addresses: *Sobell Department of Neurophysiology, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK. †Department of Psychology, Queen's University, Kingston, Ontario, Canada.
E-mail wolpert@hera.ucl.ac.uk

β -catenin/Armadillo and TCF. *Drosophila* APC has a role in maintaining adherens junctions, and also helps to tether mitotic spindles to the cortex and to orient them in the epithelial plane. In mammalian cells, APC has been implicated in cell migration. APC also helps safeguard the fidelity of chromosome segregation in mitotic cells. Loss of APC's various functions at the cell periphery may contribute to tumour progression. The multitude of these functions could explain why APC is such a potent tumour suppressor.

Is there anything we don't know? It is unclear whether APC is a genuinely multi-functional protein, or whether there is one basic molecular activity that underlies all its apparently disparate functions.

How can I find out more:

- Barth, A.I.M. and Nelson, W.J. (2002). What can humans learn from flies about adenomatous polyposis coli? *Bioessays* 24, 771–774.
- Bienz, M. (2002). The subcellular destinations of APC proteins. *Nat. Rev. Mol. Cell Biol.* 3, 328–338.
- Fodde, R., Smits, R. and Clevers, H. (2001). APC, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer* 1, 55–67.
- Kinzler, K.W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. *Cell* 87, 159–170.
- Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B. and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. *Science* 275, 1784–1787.
- Mimori-Kiyosue, Y. and Tsukita, S. (2001). Where is APC going? *J. Cell Biol.* 154, 1105–1109.
- Oshima, H., Oshima, M., Kobayashi, M., Tsutsumi, M. and Taketo, M. M. (1997). Morphological and molecular processes of polyp formation in Apc(delta716) knockout mice. *Cancer Res.* 57, 1644–1649.
- Polakis, P. (2000). Wnt signaling and cancer. *Genes Dev.* 14, 1837–1851.
- van de Wetering, M. (2002). The β -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111, 241–250.

Primer

Receptive fields

Peter Lennie

Sensory organs are filters: they use some of the information that impinges on them, and block the rest. The most aggressive filtering occurs very early. The optical components of the eye and the photoreceptors of the retina, for example, together ensure that the human visual system normally responds only to light of wavelengths from 400–700nm. Equivalent constraints confine the kinds of signals that are handled by other sensory systems. Even though much raw information is discarded by the initial transduction stage that converts visual, auditory or other sensory signals into activity in the nervous system, a tremendous amount is admitted — far more than the brain can deal with.

Consider what happens in the visual system. There are about 6 million cone photoreceptors in the human eye, each of which provides a point sample of the retinal image. But the optic nerve, through which all signals are conveyed from the eye to the brain, contains only 1.25 million nerve fibers, each of which can carry less information than a single cone. This loss of information in the eye is not

necessarily troublesome: the retinal image, like most sensory signals, is redundant, and one part of the image can to some extent be predicted from the structure of nearby parts, but for adequate prediction the right information has to be preserved.

The visual system, like other sensory systems, therefore needs selective filters designed to transmit the important structure in natural signals. Filters are embodied in individual neurons in the retina and higher stages of the visual pathway, and have characteristic properties that determine the spatial and temporal attributes of the signals they transmit. The natural way to represent the spatial attributes of the neuronal filter is through a map of the region of retina from which the neuron picks up signals. This map is called the neuron's 'receptive field'.

Figure 1 shows two representations of the receptive field of a mammalian retinal ganglion cell (the axons of ganglion cells form the optic nerve and convey the results of retinal processing to the brain). On the left, the receptive field is represented as a three-dimensional profile, with excursions above the plane denoting regions of retina where light stimulation excites the cell, and excursions below the plane denoting regions where light inhibits the cell. On the right, the receptive field is represented as a plan view overlaid on the

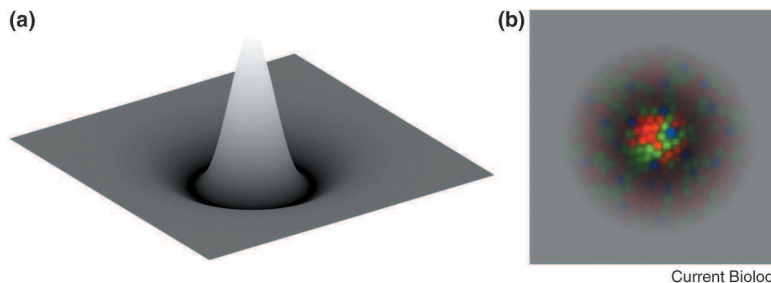
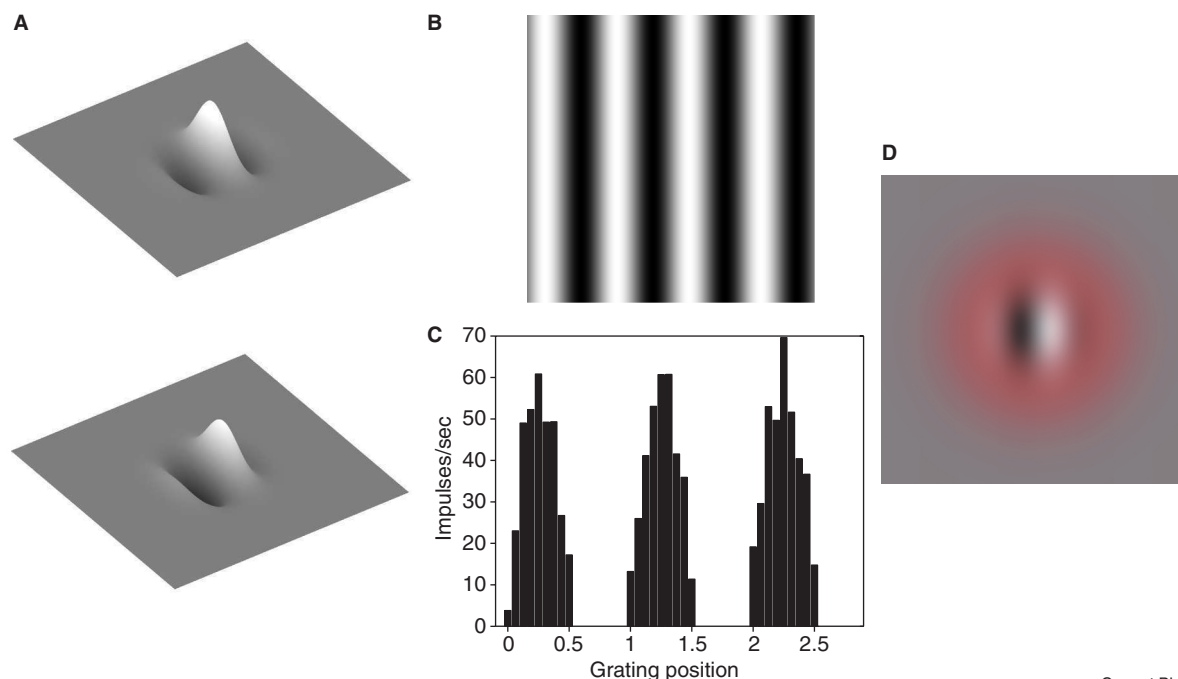


Figure 1. The receptive field of a mammalian retinal ganglion cell. (A) Three-dimensional representation of the distribution of sensitivity. The receptive field is composed of two distinct mechanisms, center and surround, each of which weights signals from photoreceptors by a two-dimensional Gaussian profile. Signals in center and surround are of opposite polarity, and are summed by the ganglion cell. The center is smaller than the surround, but the integrated sensitivities of the mechanisms are nearly the same, so the cell responds poorly to uniform illumination that covers the receptive field. (B) Plan view of the receptive field, with shading indicating the weights with which it draws on signals from underlying cone photoreceptors.



Current Biology

Figure 2. The receptive field of a 'simple' cell.

(A) The receptive field profile can be conceived as a sinusoid weighted by a two-dimensional Gaussian envelope. Different relative phases give rise to receptive fields with even (upper) or odd (lower) symmetry. (B) Sinusoidal grating patterns are often used to characterize the spatial properties of the receptive field. (C) The simple cell's response to a grating pattern moving across the receptive field is a truncated and distorted sinusoid. (D) The simple cell's response is regulated by gain-controlling signals accumulated from a region (here identified by red shading) that contains and extends beyond the receptive field.

mosaic of cone photoreceptors that pick up the image, with the lightness/darkness indicating the sign and strength of the signal received from each cone.

This kind of receptive field — often called a 'center-surround' receptive field — confers interesting properties on its cell. Because the cell generally sums the signals gathered from all points in the receptive field, it will respond poorly to uniform illumination of the whole receptive field — excitatory signals are broadly balanced by inhibitory ones. More generally, the structure of the receptive field enables the cell to block signals that are common to the center and surround, in the manner of a differential amplifier. As a result the neuron transmits information about local spatial contrast in the image.

For any visual system capable of rich analysis of images, it is important not to discard information too early. Neurons with center-surround receptive fields generally behave as simple

linear filters which impose a modest transformation on the visual image. They discard the least important information, and condition the image for economical transmission to mechanisms that undertake fuller analysis of it. They are the commonest kind in the retinas of higher mammals.

Other sorts of neurons have more elaborate receptive fields. Figure 2A shows examples of a relatively simple extension of the receptive field in Figure 1. The distribution of sensitivity in excitatory and inhibitory regions is different, and confers a greater spatial selectivity on the neuron, notably making it tuned for oriented contrast in the image. Otherwise, the neuron accumulates signals linearly, and can still be thought of as a simple image filter, constructed straightforwardly through combining inputs from appropriately arranged center-surround receptive fields. The structure of a receptive field can be explored with methods for

analyzing linear systems; sinusoidal contrast patterns are therefore often used (Figure 2B). Neurons with this kind of receptive field — called 'simple' cells — are common in mammalian primary visual cortex.

Although the cells illustrated in Figure 2 accumulate contrast signals linearly, their responses — expressed as the rates at which they discharge spikes — do not accurately track variations in this signal, principally because simple cells have little or no resting discharge, and cannot respond at rates below 0. Their responses to patterns that move across the receptive field are consequently halfwave rectified (Figure 2C).

Simple cell receptive fields are spatially localized, oriented, spatial-frequency-selective filters. For any one small region of retina, a set of simple cells tuned to a range of orientations and spatial frequencies samples the image. A recent insight about them is that the design of their receptive fields is well-suited to

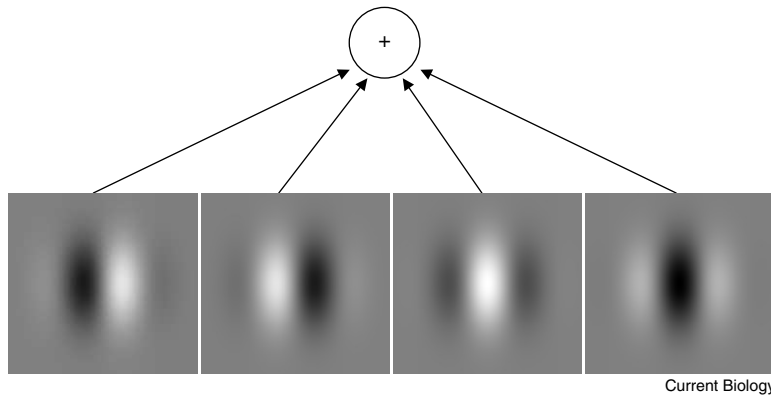


Figure 3. The receptive field of a 'complex' cell. Complex cells seem to be driven by signals from at least four simple cells, the receptive fields of which lie in the same place but prefer stimuli in different spatial phases. At least one of the simple cells will be active no matter what the spatial phase of the stimulus on the complex cell's receptive field, and because the driving signal is half-wave rectified, the complex cell will always respond.

represent the structure of natural images efficiently. That is, given the task of representing the spatial properties of natural images, and the requirement that it be accomplished by minimizing the statistical dependence among activity in neurons that respond to the image, a good choice is to use filters that have receptive fields of the kind found in simple cells.

Although such a set of linear filters helps represent the image efficiently, the complex structure of most natural scenes ensures that activity in different filters will still be substantially correlated. Real simple cells seem to possess additional, nonlinear mechanisms that can act to reduce the statistical dependence among their responses to natural images.

A simple cell has what is essentially a divisive gain control, which automatically regulates the rate at which spikes are generated, according to the contrast signal accumulated from its receptive field. This gain control is driven, not just by the individual simple cell, but by signals gathered from a pool of other cells that have receptive fields nearby. The greater the activity in this pool of neurons, the more the simple cell's gain is reduced. As the pool contains neurons that are tuned to a range of spatial frequencies and orientations and positions, the

aggregate gain controlling signal tends to be more broadly tuned than the signal that arises in the receptive field, and is accumulated from a larger region of visual field.

The gain-controlling machinery does not excite the cell directly, but as its signal is gathered from an area larger than the receptive field, part of it can be explored in the surrounding region. This 'silent surround' has been known for a long time, but only recently has it become clear how it works, and that it might be important for the efficient representation of images. The surround and the receptive field proper tend to prefer the same visual stimuli, so the surround most reduces the gain of the response when the cell's preferred stimulus covers both it and the receptive field. The structure of natural scenes is such that adjacent regions of the image are generally much alike, so the surround will help reduce the redundancy in the image representation provided by simple cells.

The image transformations undertaken by neurons up to the level of the simple cell can be conceived as providing a compressed description: one could reconstruct the visual image (albeit without full fidelity) from the signals available in simple cells. What succeeds this? We might conceive of additional layers of

transformation that remove progressively more complex forms of redundancy in the representation of natural images, ultimately giving rise to extremely efficient descriptions. These transformations might involve the construction of receptive fields that select for complex features in natural images. The notion that sensory neurons might be 'feature detectors' has a long history, rooted in the idea that some objects in the world are important and others are not, and that the job of sensory analysis is to distinguish the important from the unimportant.

Although converging lines of thought suggest the existence of receptive fields that analyze complex features, the kinds of features that are important to an organism, and help it to distinguish predator and prey or to identify a potential mate, need not be the kinds that allow efficient descriptions of images. We know too little about the higher-order structure of scenes, or about what is visually important to animals, to be able to stipulate features that receptive fields should be designed to detect. Instead, we explore the properties of receptive fields in real visual systems, and from these properties try to infer function.

The visual system does indeed possess neurons with highly-selective receptive fields — for example, cells that respond best to stimuli of a particular structure moving across their receptive fields in specified directions at particular speeds. In lower animals, sharply selective cells can be prominent even in the retina, but in higher mammals they are seldom seen at a lower level than the cortex, reflecting the general principle that the more complex the sensory analysis of which an animal is capable, the more likely it is to be deferred to cortex.

Some complexity is present even in primary visual cortex. Figure 3 shows the underlying structure of perhaps the commonest type of receptive field in the primary cortex, whose organization can only be inferred,

not mapped directly. It can be conceived as being assembled from the receptive fields of several simple cells with receptive fields in the same place. Each contributing simple cell prefers the image in a different position on its receptive field, so the neuron that accumulates rectified signals from the set of simple cells will have the spatial selectivity of a simple cell, but will be excited no matter what the position of the preferred stimulus within its receptive field. This kind of neuron is called a 'complex cell'.

Although only a small step away from a simple cell, the complex cell has distinctively different properties that highlight important general issues. First, it is not obvious what image features the complex cell is designed to detect. It responds to contrast power in the image, in certain orientations and within a certain band of spatial frequencies; its receptive field is not well suited to identifying edges, or the sorts of image structure we normally think of as a feature. It is better suited to conveying information about the general structure of surfaces, as coarse or fine-textured, or having characteristic material properties. The complex cell's sensitivity to contrast power also makes it potentially useful in encoding image movement in contexts — such as optic flow arising from self-motion — where it might not be important to know what in the image is moving. Similarly, the complex cell can be well-suited to encoding position in depth, providing a signal about where an object is without necessarily indicating what the object is. Second, in discovering particular kinds of structure in the image, complex cells irretrievably discard information about its other attributes. To the extent that these discarded attributes are important for the overall representation of the image, parallel machinery must exist to detect them. We know generally how this machinery is organized, though the details are enigmatic.

In the first visual cortical area, where simple cells and complex

cells arise, there are few signs of other major types of cell that might be selective for more specialized image features. Indeed, to the extent that other distinct kinds of cell can be identified at all in primary visual cortex, they seem to be organized for handling information about color, and have receptive fields that are less selective than those of simple cells. Detectors for more specialized image features must emerge in higher cortical areas to which primary cortex transmits information, and their receptive fields must be constructed from signals delivered by simple cells and complex cells.

The gross organization of visual cortical areas and relationships among them are well established, principally through anatomical work. Each holds a two-dimensional map of the retinal surface. The map is distorted, though preserves the topology. The multiple areas are organized hierarchically with generally more than one at each level, so as we ascend the hierarchy we might expect to encounter neurons with receptive fields of increasing specificity and complexity. It has been surprisingly difficult to evaluate this idea.

The physiologist who explores receptive fields in higher visual areas has little guidance about what to expect. As a result, a receptive field is generally characterized using one of two approaches: by delivering standard stimuli, such as sinusoidal gratings, which under the right conditions permit powerful inferences about the underlying receptive field structure; or by using a repertory of perhaps richly structured patterns, some of which might contain the feature(s) to which the receptive field is tuned. Both approaches suggest that neurons in higher areas have more exacting stimulus requirements than those in early cortex, but neither approach has yet revealed much about the essential differences among areas or the

principles used to analyze the image.

As we learn more about the statistical structure of the visual world, and about the aspects of it that are most important to different animals, we can expect more sharply guided physiological explorations to reveal the details of the successive stages of image analysis.

Acknowledgement

I am grateful to Jonathan Peirce for help with figures.

Key references

- Enroth-Cugell, C. and Robson, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J. Physiol. (Lond.)* 187, 517–552.
- Hubel, D.H. and Wiesel, T.N. (1962). Receptive fields, binocular interactions and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* 160, 106–154.
- Ringach, D.L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* 88, 455–463.
- Simoncelli, E.P. and Olshausen, B.A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.
- Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* 9, 181–197.
- Cavanaugh, J.R., Bair, W. and Movshon, J.A. (2002). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *J. Neurophysiol.* 88, 2530–2546.
- Schwartz, O. and Simoncelli, E.P. (2001). Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4, 819–825.
- Barlow, H.B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception* 1, 371–394.
- Movshon, J.A., Thompson, I.D. and Tolhurst, D.J. (1978). Receptive field organization of complex cells in the cat's striate cortex. *J. Physiol. (Lond.)* 283, 79–99.
- Lennie, P. (1998). Single units and visual cortical organization. *Perception* 27, 889–935.
- Gallant, J., Braun, J. and Van Essen, D.C. (1993). Selectivity for polar, hyperbolic and Cartesian gratings in macaque visual cortex. *Science* 259, 100–103.

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.

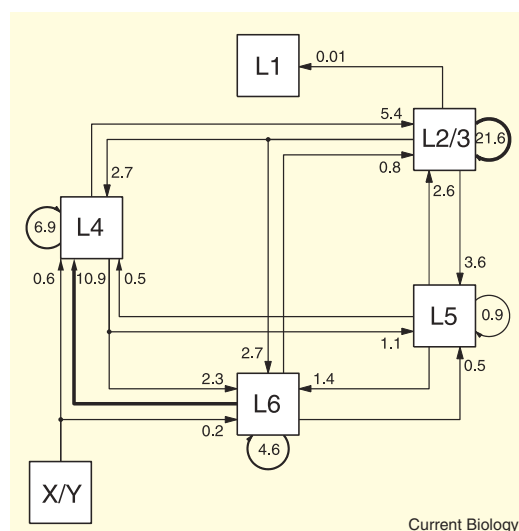
One clue is the dominance of local cortical synapses over those provided by individual afferents of a given cortical area, as shown in Figure 1. Overall, the vast majority of cortical excitatory synapses and virtually all inhibitory synapses originate from neurons within cortex and most of these synapses 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).



Current Biology

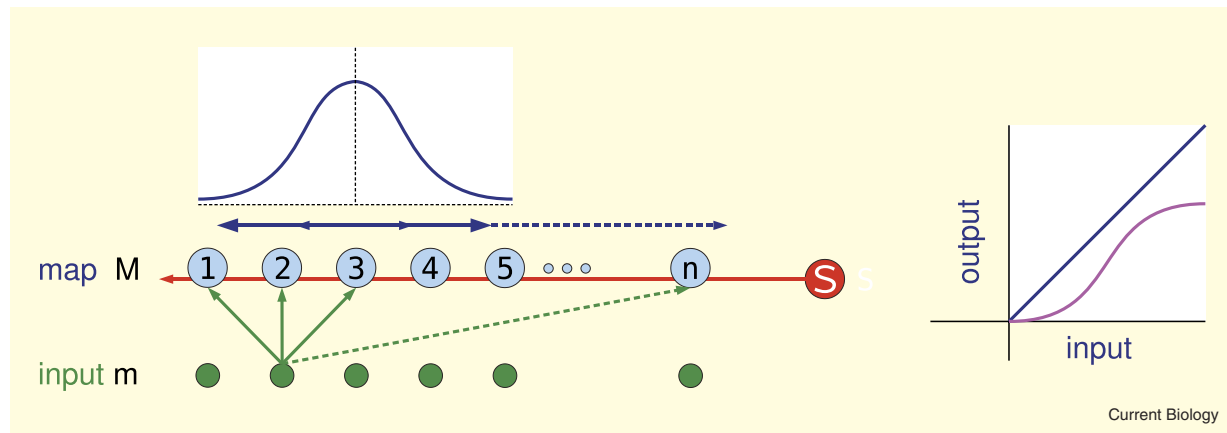


Figure 2. Simple model of recurrently connected neurons with rate coded outputs.

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 through 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

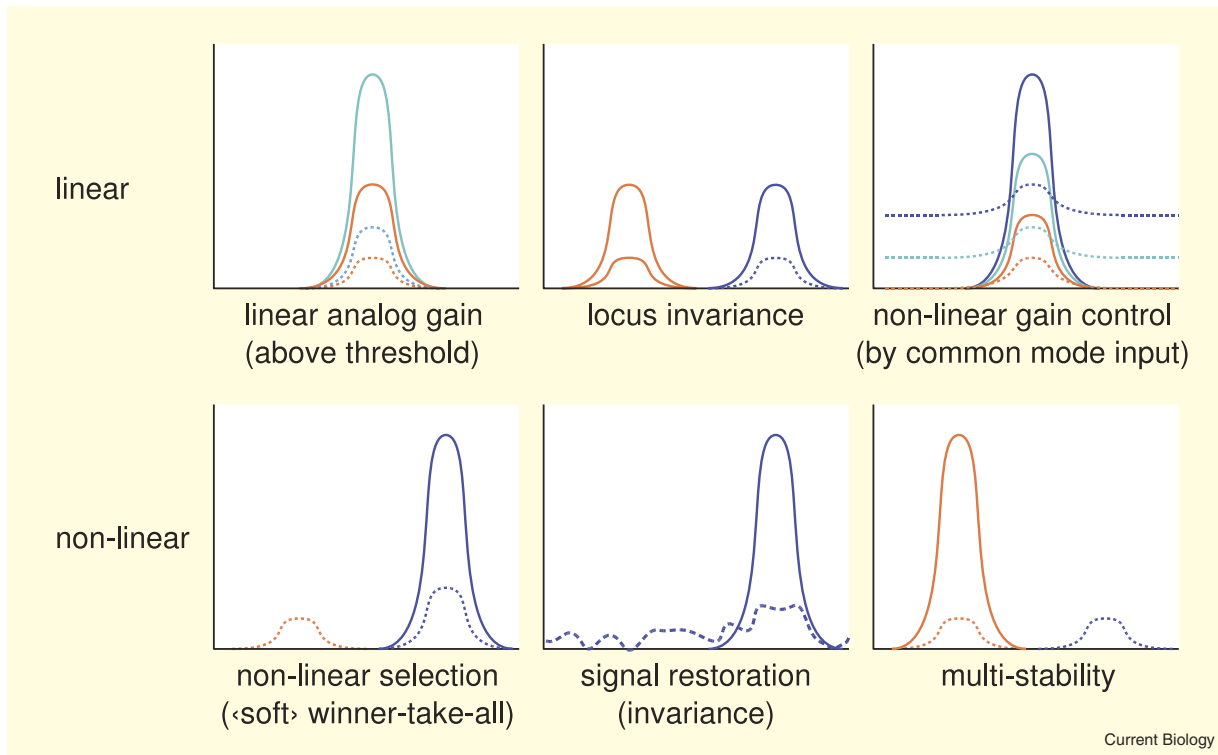


Figure 3. Six interesting functional properties of the recurrent network described in Figure 2.

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

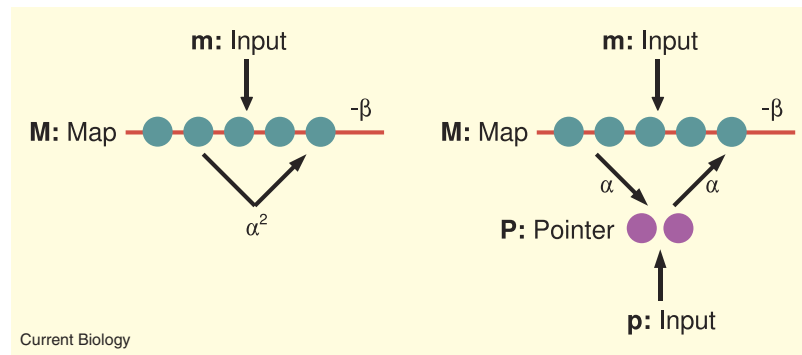


Figure 4. Comparison of standard recurrently connected network left (equivalent to Figure 2), and the ‘pointer-map’ configuration of recurrence.

Both networks have inhibitory feedback ($-\beta$, 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.

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.

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. Nathan
Wiley, 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

larger the better — is absolutely vital. Without sentries, meerkats cannot feed safely; without helpers, they cannot breed successfully; without strength in numbers, they cannot hold territory. Through cooperation, individuals can increase the size and success of their group, and in so doing, improve their own individual prospects for survival and reproduction.

Do all meerkats help equally?

No, research has shown that the amount of help provided typically depends on the ability to help: well-fed individuals are more likely to expend energy for the benefit of others. But this is not to say that meerkats simply help indiscriminately. Female helpers, for example, preferentially feed female pups — probably because these pups, being the philopatric sex, represent the future workforce that might one day help to raise the helpers' own litters.

How is it that we know so much about meerkats?

Meerkats' striking sociability and their tendency to become habituated to the presence of human observers, make them eminently tractable model organisms for investigations into the evolution of social behaviour. A large-scale, long-term field study of meerkats in the southern Kalahari was initiated in 1993 by Tim Clutton-Brock of the University of Cambridge. This project, involving hundreds of thousands of man-hours of detailed behavioural observations coupled with long-term pedigree and life-history information has provided a wealth of data on cooperative breeding behaviour and constitutes the best database of such information that exists today for social mammals.

What else have we learned from meerkats?

Apart from serving as a model for the evolution of cooperation, the Kalahari meerkats have provided unique opportunities for research in other areas of biology too. Through observation and experiments, Marta Manser and her group at the University of Zurich have 'decoded'

many of the 30 or so calls, and the numerous postures, displays and olfactory signals that meerkats use. This work extends our understanding of animal communication and cognition. For example, meerkat alarm calls, which categorise threats according to both class and urgency, exhibit a level of complexity more typically associated with humans. Other behavioural studies have used meerkats to investigate why animals play, and recently they were used to neatly demonstrate that animals can, and do, actively teach their young. On the ecological front, meanwhile, meerkats have provided insight into the demographics, dispersal, and population genetics of social carnivores. For instance, long-term population data clearly demonstrate the "Allee effect": if meerkat group sizes fall too low, recruitment tails off rapidly and local population crashes can ensue. Finally, at a physiological level, meerkats are helping us to better understand the energetics, endocrinology and thermoregulation of small desert-adapted mammals.

Where can I learn more about meerkats?

- Clutton-Brock, T.H., O'Riain, M.J., Brotherton, P.N.M., Gaynor, D., Kansky, R., Griffin, A.S., and Manser, M. (1999). Selfish sentinels in cooperative mammals. *Science* 284, 1640–1644.
- Clutton-Brock, T.H., Brotherton, P.N.M., Russell, A.F., O'Riain, M.J., Gaynor, D., Kansky, R., Griffin, A., Manser, M., Sharpe, L., McIlrath, G.M., *et al.* (2001). Cooperation, control and concession in meerkat groups. *Science* 291, 478–481.
- Clutton-Brock, T.H., Russell, A.F., Sharpe, L.L., Brotherton, P.N.M., McIlrath, G.M., White, S., and Cameron, E.Z. (2001). Effects of helpers on juvenile development and survival in meerkats. *Science* 293, 2446–2449.
- Clutton-Brock, T.H., Russell, A.F., Sharpe, L.L., Young, A.J., Balmforth, Z., and McIlrath, G.M. (2002). Evolution and development of sex difference in cooperative behaviour in meerkats. *Science* 297, 253–256.
- Russell, A.F., Young, A.J., Spong, G., Jordan, N.R., and Clutton-Brock, T.H. (2007). Helpers increase the reproductive potential of offspring in cooperative meerkats. *Proc. R. Soc. Lond. B* 274, 513–520.

Institute of Evolutionary Biology,
University of Edinburgh, Ashworth
Laboratories, Kings Buildings, West
Mains Road, Edinburgh EH9 3JT, UK.
E-mail: a.ross-gillespie@ed.ac.uk

Primer

Structure and function of the cerebral cortex

Stewart Shipp

The grey matter of the cerebral cortex is a convoluted, layered sheet of tissue, 2–3 millimetres thick in man but with a surface area of several hundred square centimetres. This is not an adaptation to promote gaseous exchange, or heat loss — rather, if the grey matter is compact in at least one dimension, it is outgoing axons that may readily escape it; once outside, they club together and form the cortical white matter. If grey and white were intermixed, the average separation of neurons would be greater, creating extra neural 'wiring'. The speed of cortical computation would suffer accordingly.

The principle of economic wiring can also be invoked to account for regional specialisation of function across the surface area of the cortex. Put simply, neurons performing similar roles need to communicate, and do so more efficiently if nearby. When microelectrodes were first used for cortical recordings, an immediate discovery was that neurons in a radial column — a column perpendicular to the plane of layering — shared similar response properties: their receptive fields might all be located on the same patch of skin, or be selective for the same orientation of a contour at a certain location in the image on the retina. Moving tangentially through the sheet (parallel with the plane of layering) the discovery was that neighbouring columns have neighbouring receptive fields — the ensemble of columns ultimately giving rise to a cortical map of the relevant sensory surface. In sensory cortex, this engenders the 'one map, one area' principle for parcelling the cortical surface into discrete areas, each of which is thought to have some nuance of functional

specialisation. Cortical areas are richly interconnected — with each other and with subcortical structures — and the layering of the cortex reflects the radial organisation of all these input–output relationships. Indeed, the layered pattern is rather uniform over the expanse of the sheet, as if to serve basic ‘housekeeping’ operations that generalise across cortical applications as diverse as colour vision, speech and music.

One might assume that, if outputs to different destinations arise from separate levels in a column, there should be some differentiation in function to match. Thus, in principle, the study of cortical layers is the study of cortical specialisation in the radial dimension, and how it is conserved from area to area. Of course, there are subtle variations in the laminar structure across the cortical sheet, which form a complementary means of identifying separate areas, as for example in Brodmann’s century-old scheme for human cortex. Sometimes the laminar variations are not so subtle — the contrast between primary motor and primary visual areas (M1 and V1) being the most emphatic (V1 showing the highest degree of laminar differentiation, and M1 the least). These variations in cortical architecture have long been treated purely cartographically, betraying a lack of any analytic insight into the way different applications might modulate layer structure and function. This is largely because, as documented below, our appreciation of layers is still rooted rather more securely in anatomical than physiological cortical characteristics.

Cellular machinery

Neurons come in two main forms: excitatory (pyramidal) and inhibitory. Pyramidal neurons are named for their prominent apical dendrite, which typically points superficially. Customarily, a neuron ‘belongs’ to the layer in which its cell body is sited — even if the apical and basal dendrites, between them, span several more layers, picking up a broader range of signals. Pyramidal

neuron dendrites are covered in spines (specialised postsynaptic structures) and the density of spines, coupled to the degree of dendritic branching within a layer, indicate the parent neuron’s commitment to sample signals from that layer.

Inhibitory neurons, despite being in a minority (20%), are rather more diverse in their morphology. Some forms have attracted familiar names, for example basket cells and chandelier cells, on account of their characteristic axonal ramifications, forming ‘baskets’ around cell bodies, or making multiple strings of contacts (‘chandeliers’) around axon initial segments. The names are not simply fanciful, because the contact region on the target cell — axon, cell body, proximal or distal dendrites — is a useful component of a systematic classification. Inhibitory neurons are also known as local circuit neurons, or simply interneurons, because, in the cortex at least, their axons are purely ‘intrinsic’ — they do not enter white matter and make only short-range, local connections. Some pyramidal cells act similarly, but others make both short and long range connections. The latter are carried by axons passing through white matter termed ‘extrinsic’.

In a classic of the developmental literature, the cortex is created ‘inside out’, as newborn pyramidal cells migrate, radially, away from the proliferative zone to form layer 6 first, and layer 2 last. Layer 1, which in adults is largely cell free, originates from the marginal zone of the developing cortical plate, a likely source of the signalling molecule, semaphorin 3A, known to polarise the initial growth of apical dendrites. Inhibitory cells have diverse origins, migrate tangentially, and at least one major class displays an ‘outside-in’ gradient of neurogenesis. The genetic mechanisms controlling this process are beginning to be unravelled; for instance, different sets of transcription factors are found to regulate the generation of neurons destined for layers 6 and 5, and those later destined for layers 4, 3 and 2. As we shall see, there are several sets of extrinsic

connections that make a similar distinction.

How cortex connects to itself

The diagram in [Figure 1](#) is billed as the connectivity of a generic, primate, non-primary visual area (also incorporating knowledge of rat and cat sensory cortices). A perverse choice, perhaps, given that details are scarce and there is, without exaggeration, an order of magnitude more information available concerning the intrinsic connectivity of primate V1 than for any other visual area. Yet V1 is just about the most exquisitely laminar structure known to neurobiology, likely to possess many idiosyncratic features. The detection of generic principles requires a broader cortical survey.

We can make a start following the well-beaten tracks of cortico-cortical communication. [Figure 2](#) depicts a string of areas forming a serial pathway. Looking from left to right, the pathways arriving at an area carrying new sensory signals terminate in layer 4; there is then an internal relay to layer 3, which is the primary source of output to layer 4 of the next area in line. The sequence is initiated by input into layer 4 of the primary cortical area from a primary (or ‘first-order’) relay nucleus of the thalamus, and is known as an ascending pathway. There is also, typically, a descending (or feedback) pathway issuing from layers 6 and 5, and terminating in layers 6, 5 and 1 of the area below.

The ascending-descending (A–D) patterns are sufficiently regular to constitute ‘rules’ of cortical connectivity: for instance, there are no known instances where reciprocal connections between a pair of areas are both ‘A’ (or both ‘D’) in format. Hence the laminar patterns can be used to infer the relative hierarchical status of a pair of connected areas, even where the relative distances from the primary area are unknown. Ultimately, a hierarchical chart encompassing the entire system can be constructed with all the constituent areas allocated to a succession of tiers. Higher tiers may involve several areas interconnecting with a symmetrical, undifferentiated laminar pattern,

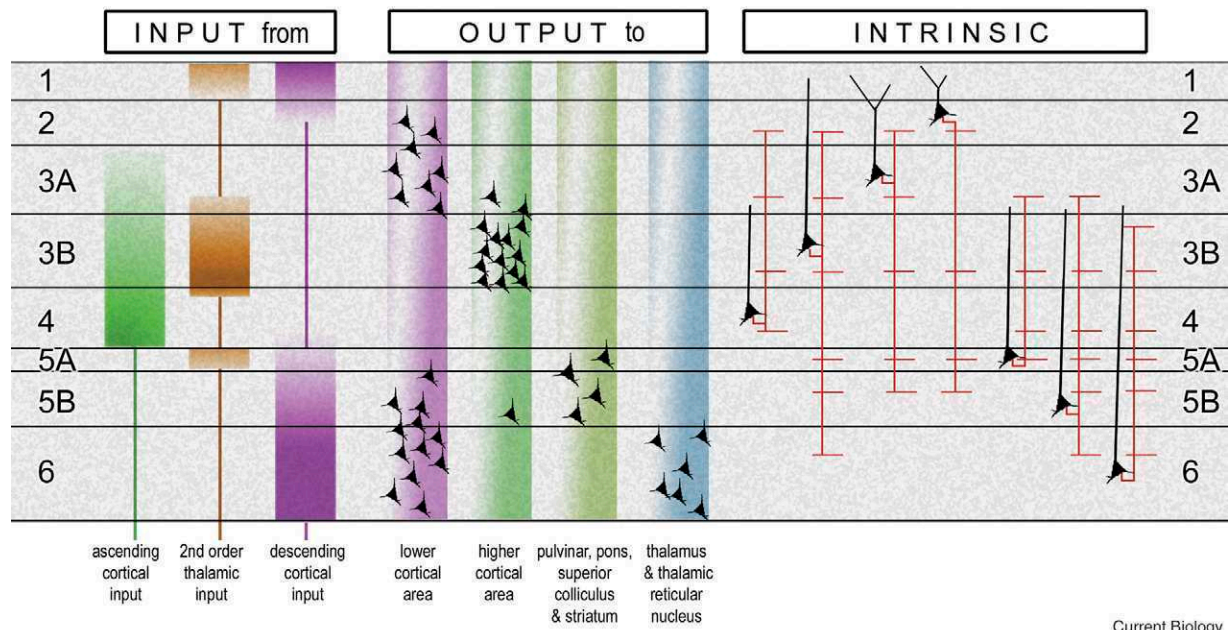


Figure 1. Summary of inputs, outputs and intrinsic excitatory connections of a generic, nonprimary visual area of primate cerebral cortex.

referred to as a 'lateral' connection. The exercise can be conducted for each of the visual, auditory and somatosensory systems and, to a degree, there are similar laminar patterns of connectivity within non-sensory frontal cortices.

There is, naturally enough, a rather greater degree of variability in laminar connectivity than the A-D scheme (as in Figure 2) admits. First off, connections can skip one or more levels in the hierarchy, and in so doing obey a 'distance rule': the connection actually adheres more tightly to the A-D laminar pattern the more levels it traverses. Connections between areas at immediately adjacent levels tend to be a little more liberal. For instance the ascending connection may include a component arising from neurons in layers 5 and 6, and it may terminate in layer 3 as well as layer 4. The descending connection may include a contribution from neurons in layers 2 and 3A, and terminate in a way that is less polarised toward the outer layers, 1 and 6. Second, where distinguishing A or D from L patterns becomes tricky, the laminar terminations are often more informative than the cellular origins. Third, where individual axons of a given connection are reconstructed in detail, there are

typically some details apparently unique to that connection (given that, to date, only a few connections have been scrutinised so closely). Overall, however, the systematics are unquestionably regular enough to begin to enquire how the A-D pathways work, and how they differ from each other.

The ascending pathway

The primary target of the ascending pathway, layer 4, is defined cytoarchitecturally by the dense packing of small cells and is also known as the 'granular layer'. Layer 4 cells in turn relay signals to layers 3 and 2. However, there is also a component of the ascending pathway that terminates directly within layer 3B, the primary source of the ascending output to the next stage. Does this mean that there is, minimally, a monosynaptic relay through each area? A shrewd guess is that this not so, although this is only known definitively for area V1, where the minimal relay (of thalamic signals) is disynaptic. The purpose of the pathway is to process rather than simply 'relay' signals, which requires more elaborate synaptic interactions.

The granule cells of layer 4 include a minority of inhibitory types, whose postsynaptic

structures can be recognised by electron microscopy, forming about 10–20% of all contacts made by the ascending axon terminals. Most of these terminals contact the dendritic spines of excitatory cells; however, fewer than 5% of excitatory synapses with spines in layer 4 are actually formed by the ascending input. The remainder are intrinsic connections between layer 4 pyramidal cells, and this may be a means of amplifying the incoming signal.

There are also extensive lateral connections formed between pyramidal cells in layers 3 and 2. Typically these are patchy in appearance, as if nests or clusters of cells with certain properties in common — such as similar orientation tuning, in visual cortex — selectively connect with each other. There is, in addition, the bewildering diversity of selective connections made by inhibitory neurons to ponder. It is reasonable to suspect that certain basic neural operations, known in shorthand as a 'canonical microcircuit', are repeated over and over again in the serial chain of areas — ultimately resulting in the synthesis of such exemplars as face selective neurons — but the details are little understood.

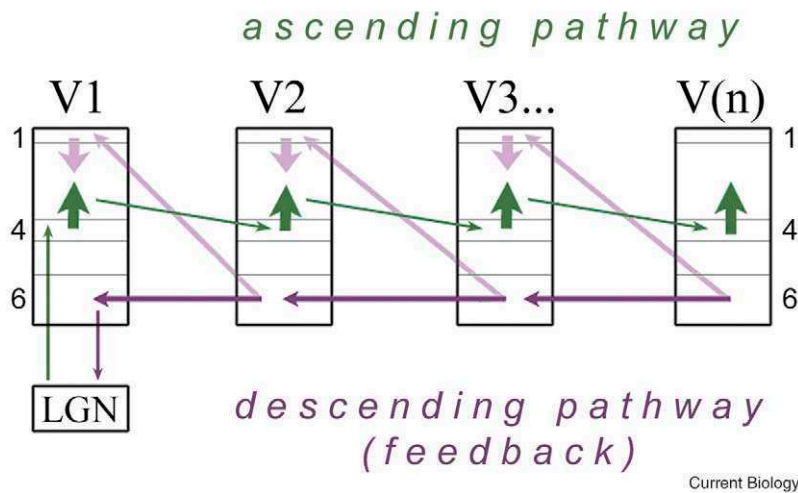


Figure 2. The laminar characteristics of ascending and descending connections along a serial (hierarchical) pathway of sensory areas in primate cortex.

Although layer 4 is a distinct cytoarchitectural structure, it is valid to question how far it, or any other layer, acts as a discrete functional sub-unit, as there are so many ways in which both excitatory and inhibitory neurons show total disregard for layer boundaries. For example, layer 4 contains spiny apical dendrites of layer 5 pyramids, and spiny basal dendrites of layer 3 pyramids; furthermore, the apical dendrites of layer 4 pyramidal neurons pass into layer 3. A valid functional subcompartmentation should show some systematic discontinuities in connectivity across layer boundaries, and these can be hard to identify. One good example is the absence of excitatory, intrinsic feedback from layer 3 to layer 4. Not only do the pyramidal cells of layer 3 fail to make synaptic contacts as their axons pass through layer 4 toward white matter, but their collateral axons ramifying within layer 3 also avoid contact with the apical dendrites of layer 4 pyramidal neurons. This clearly effects some degree of separability between neural operations within the two layers, consistent with a minimal 'two-stage' processing within each tier of the ascending cortical pathway.

The descending pathway

As mentioned above, both the origins and terminations of the descending pathway tend to

avoid the middle layers. Layer 1, often the principal target of this pathway, is a unique stratum almost entirely devoid of neuronal cell bodies, apart from a few inhibitory neurons. It is composed largely of incoming axonal ramifications and the branching apical dendrites of pyramidal cells, mostly those of the upper layers but some sited as deep as layer 6. The remaining component is a rich contingent of glial cells, perhaps signifying a distinct local blend of neurochemistry, but one whose physiological significance remains wholly mysterious.

Sourcing the origins of the descending pathway, the major inputs to layer 5 come from layers 2/3, often formed by extensive collateral branches of output axons heading toward white matter. The input from layer 2/3 to layer 6 is, by comparison, rather weaker, and more restricted to underlying neurons in the same radial column. Layers 5 and 6 themselves are interconnected but, of the two, layer 6 appears less intensively integrated into the local intrinsic circuitry; conversely, layer 6 is typically a richer source and target zone for the extrinsic circuitry of the descending pathway.

The action of the descending pathway is often termed 'feedback', an implicit assumption that it is triggered as an echo of ascending signals. However, the simple overview of A-D systematics

(Figure 2) immediately suggests that each pathway, using a distinct set of layers, is capable of some autonomy. Anatomically, an independent chain of activation relayed backwards from layer 6 to layer 6 looks possible. Or, in a cognitive context, the descending pathway could serve as a conduit for top-down effects related to attention. Using the visual system as an example, there are many indications that attention to particular objects, features or locations in the visual scene is associated with increased activity of the neurons with matching receptive field characteristics. In a similar vein, the descending pathway could mediate expectations, or higher level hypotheses as to what may be present in the retinal image (imagine looking for a familiar face in a crowd). The intrinsic interactions across layers could then be viewed as a matching process, between visual signals looking for an explanation and visual hypotheses searching for evidence.

The fact that this is couched in semi-theatrical terms reveals a paucity of any more profound mechanistic understanding of the interaction between these two streams. One proviso to the above scheme is that descending connections are generally considered to be 'modulatory' rather than 'driving' in action — capable of modifying ascending activity, but not of initiating activity *de novo*. The ascending pathway appears to have a driving character, as illustrated by the systematic growth in receptive field size caused by spatially convergent connections from one level to the next. The neurons at each level are capable of being activated by a small subset of their afferents. There is at least as much, if not more, spatial convergence in the descending pathway, and if the synaptic efficacy were also no different, receptive fields at the lower level would soon inherit the pooled size of the higher level fields. As this is not observed — in any layer of any area of any species so far studied — it is relatively safe to conclude that the synaptology of the ascending and descending pathways must differ significantly.

The strict implication is that descending activity cannot normally propagate across levels in the absence of an ascending substrate with which to interact. But if, in abnormal circumstances, it were to have this capacity, we might be looking at a hallucination.

Variants of architecture

If the A-D system and its proposals for layer specialisation of function has any claim to provide a general account of the cerebral cortex, it must survive a stiff challenge from the motor system. The root of the problem is that motor cortex is, rather famously, agranular, and the lack of a layer 4 hinders the systematic diagnosis of A-D laminar patterns. If seeking to explore an analogy with sensory cortex, it is also unclear, *a priori*, which direction along the motor pathway is 'ascending': is it (a) from primary motor cortex (M1) to premotor cortex, on the grounds that M1 is closer to the effectors (spinal motoneurons), and has the simpler motor fields; or (b) the opposite, on the grounds that the higher premotor centres should 'drive' the motor commands issued by M1?

In terms of anatomy, model (a) is more correct, in that connections from higher to lower premotor areas, and to M1, show identifiably 'D' patterns of laminar connectivity. It is, however, difficult to find any good examples of a classical 'A' pattern of connection between motor areas. Instead, connections from M1 to premotor areas target all layers roughly equally (in the manner of a lateral connection in sensory cortex). Evidently, the lack of a clear ascending system relayed through a granular layer 4 might reflect the lack of a stream of input data that requires serial re-processing in a way analogous to sensory cortex. But how, then, can the full A-D model be adapted to the operation of motor cortex?

There are, in fact, ascending cortical inputs to motor areas — if defined, anatomically, by terminating in the middle layers — and these derive from bimodal visual/somatosensory regions of cortex. The sensory processing is thought to convert external object locations into body coordinates (object

location as specified in terms of shoulder, elbow and wrist joint angles required to grasp that object). The sight of an object by itself is sufficient to drive some premotor neurons — as if to code a potential act of grasping that is not necessarily executed. Another interesting category of premotor activity is demonstrated by 'mirror neurons' that can be activated by the sight of an action performed by another individual, when it is congruent with the neuron's own specificity in the production of body movement.

The premotor activity could thus be thought to represent motor concepts that might be translated into action by a descending or modulatory influence upon M1. It makes sense that this should not be a deterministic, driving action upon M1, because the appropriate use of muscles will depend on the current body posture and limb dynamics. Hence, the evolution of the motor output signal might depend on another kind of distributed matching process — between external object and internal body-related activity filtering through the intrinsic circuitry of motor cortex. This permits a provisional common framework for considering cortical communications within both the sensory and motor systems.

How cortex connects subcortically

The output from motor cortex to spinal motor neurons derives entirely from layer 5 pyramidal cells. In fact, layer 5 qualifies as the 'motor' output layer over the entire cortex, because nearly all signals directing behaviour emanate from this layer. Corticospinal neurons in layer 5 may be densest in M1 but are distributed amongst several somato-motor areas and in visual cortex, for instance, layer 5 is the source of output to the superior colliculus, a midbrain centre governing head and eye movements. The same layer of frontal eye field, meanwhile, has direct output to brainstem oculomotor centres. Other targets of layer 5 include the cerebellum (via the brainstem pontine nuclei), the striatum and the thalamus.

As mentioned above, layer 5 is closely integrated with layers 2 and

3, as if to sample the activity of these superficial layers. Generally speaking, layer 5 neurons have relatively large receptive fields, and are 'trigger-happy' — their high spontaneous activity reflects a resting membrane potential that is relatively close to the threshold for initiation of an action potential. The best-documented example is the output from V1 to superior colliculus, which is found to arise from large pyramidal neurons with apical dendrites rising to layer 1. Compared to other layers, or even other layer 5 neurons, these cells are less selective for the orientation and direction of a visual stimulus and totally insensitive to its colour. These are rational characteristics for an output signal that basically serves to provide spatial coordinates for the acquisition of target objects by the oculomotor system. A sophisticated target description (apart from basic details concerning location/size/shape) is, presumably, irrelevant to the ocular or skeletomotor apparatus, and can be locked up in cortical circuitry for the unique appreciation of our cognitive faculties.

Layer 6 also connects to the thalamus (including its inhibitory sheath, the thalamic reticular nucleus) — but as this is merely the outgoing component of a cortico-thalamo-cortical loop, layer 6 does not act as an alternative gateway for cortical signals to directly influence behaviour. As a rule, neurons making either cortical or subcortical connections form separate populations. There is considerable diversity within each population, although the picture is confused by the fact that the outputs are frequently carried by axons bifurcating toward two or more targets: much or all of the layer 5 output to thalamus, for instance, is formed by collaterals of axons travelling toward the brainstem. There is one prominent set of neurons, the giant Meynert cells of primate primary visual cortex, that manage to flout all the above generalisations, in that they: are sited in layer 6; connect, via bifurcating axons, to both the superior colliculus and a higher cortical area, V5/MT; and have short, stubby apical dendrites.

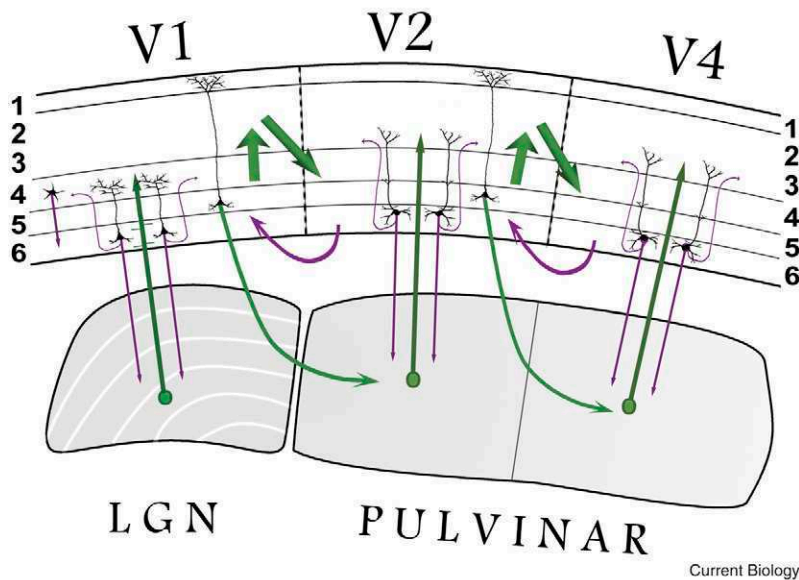


Figure 3. Schematic treatment of the driving and modulatory interactions between visual cortex and visual thalamus.

Green connections are inferred to be driving, in the sense of being capable of initiating activity *de novo*. Modulatory connections (purple) are thought not to have this capacity. The topographic arrangement of layer 5 and layer 6 afferents in pulvinar is not known to be quite so precise as illustrated here (specific territories governed by specific V2/V4 afferents). Also omitted are known geniculate afferents to V2 and V4.

How do layers 5 and 6 differ?

A comparison of the respective thalamic connections is highly instructive in regard to the differing functional roles of the subcortical outputs issuing from layers 5 and 6. The former is regarded as 'driving' and the latter as 'modulatory', as mooted above for the action of the A and D cortical pathways, although the thalamic evidence is the more robust. These differing characteristics are part and parcel of the distinction between first and second order thalamic nuclei, as shown in Figure 3. A first order nucleus is a classic 'relay' nucleus, such as the dorsal lateral geniculate nucleus (LGN) of the visual system, which feeds V1 and receives feedback from layer 6 of V1. Output from layer 5 of V1 does not terminate in the LGN, but only in the second order nucleus (i.e. the pulvinar nucleus within the visual system).

The second order nucleus is defined by the fact that it receives scant direct sensory input; instead, its driving inputs are provided from cortical layer 5. Anatomically, layer 5 terminals in the pulvinar match retinal terminals in the LGN. Each forms only a small minority of the

overall input to their respective nuclei, but the terminals are large and strategically placed on proximal dendrites (close to the cell body). It is the retinal terminals that confer their response characteristics upon geniculate cells, and layer 5 afferents to pulvinar are expected to achieve a similar driving effect. By comparison, terminals of layer 6 afferents are smaller, and terminate less focally.

The interaction between V1 and the LGN has been studied more intensively than any other feedback system. Cortical feedback appears to modulate the response of the geniculate neuron to a feature according to its context within the surrounding image. On a blank surround, the response is amplified (positive feedback). One such, notable, effect is a capacity to synchronise the responses of nearby LGN neurons responding to the same feature, such as an extended contour.

Within V1, there is a uniquely close association between layers 4 and 6 (which can be seen across species, and in other primary sensory areas too). Evidence for this association is that geniculate afferents terminate in layer 6 as

well as layer 4; there are direct outputs from layer 4 to layer 6, which are more prominent in primary cortices than elsewhere; the layer 6 pyramidal cells sending output to the LGN also sample layer 4 activity via branches of their apical dendrite within layer 4; and these same neurons also have ascending axon collaterals that arborise specifically within layer 4 (Figure 3). Anatomically, the 6-to-4 loop thus resembles a short-circuit amidst the cortico-geniculate feedback system, and there are reasons for thinking that it, too, is modulatory (it terminates mainly upon inhibitory neurons and, like monocularly driven LGN neurons, layer 4 neurons do not inherit layer 6 properties such as binocularity).

Corticogeniculate layer 6 neurons, and other layer 6 neurons (which may make intrinsic or extrinsic cortical connections) differ in some key respects. The latter have relatively widespread local axonal connections, contacting mainly other pyramidal neurons; the former (corticogeniculate) neurons have a more locally restricted axonal territory and a preference for contacting nearby inhibitory interneurons. The corticogeniculate neurons are therefore implicated in a local competition, mediated through the local inhibitory network. And, if competing with each other, this will effectively be a contest for the representation of a particular feature, for example between the orientations represented in adjacent cortical columns.

The functions of subcortical loops

Assembling the various pieces of this jigsaw, the picture is that of a corticogeniculate feedback system mediating a competitive process between the alternative features that might be represented at a particular image location, acting to select and amplify the dominant feature during any given period of eye fixation. Can such an interaction between V1 and LGN be used as a general model for the interaction between non-primary cortex and second-order thalamus? Given the lack of experimental data, this has to remain an open question. The obvious puzzle is that higher areas (nominally 'V2' and 'V4' in Figure 3) have a dual source

of likely driving inputs: from layer 3 of a lower area, relaying directly to layer 4, and from layer 5 of the same lower area relaying, via the thalamus, mainly to layer 3.

It is normally assumed that the hierarchical development of sensory representations is primarily constructed by the direct, ascending cortical pathways that represent a wider range of features at a higher level of precision. The second-order thalamic input should therefore adopt a different kind of role, and one reasonable inference is that it plays a part in spatial selective attention. Again, using the visual system for illustration, there is a well-documented anatomical and psychophysical overlap between the control of eye-movements and the deployment of covert attention, a consideration which helps to make sense of the fact that the layer 5 outputs to the pulvinar originate from neurons also communicating with the superior colliculus. Hence, assuming that the layer 5 outputs carry an object-selective signal, they could act via the pulvinar to exert a regulatory influence over the transcortical networks formed by the information-rich, superficial layers — specifically, perhaps, to propagate object-selective bias between the different kinds of object descriptions found in different visual areas.

The subsequent reciprocal interaction between layer 6 and the pulvinar could be more analogous to the corticogeniculate system, at least insofar as many layer 6 neurons have apical dendrites and axonal collaterals arborising within the pulvinar terminal zone (layer 3, as opposed to layer 4). The neural dynamics of the interaction are certainly unknown but, commensurate with an attentional process, it is at least worth noting that the topography of cortico-pulvinar relationships is considerably less precise than that of corticogeniculate relationships, permitting neural competition to develop over broader stretches of cortical maps, and hence between objects at separate, distinct spatial locations.

Epilogue

The complexities of cortical circuitry are nothing short of fiendish, and

the problem of integrating genetic, morphological and physiological details from diverse cortical areas and across diverse species is a worthy challenge to the burgeoning science of neuroinformatics. Though inconsistencies abound, the fact that some trans-areal, trans-specific generalisations are possible, and justified, is a quite remarkable observation. Following the strategy of 'know thine enemy', it appears that the cortical fiend has some interesting habits, which we can usefully begin to tag with some shorthand, functional labels.

Further reading

- 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.
- Barone, P., Batardiere, A., Knoblauch, K., and Kennedy, H. (2000). Laminar distribution of neurons in extrastriate areas projecting to visual areas V1 and V4 correlates with the hierarchical rank and indicates the operation of a distance rule. *J. Neurosci.* 20, 3263–3268.
- Douglas, R.J., and Martin, K.A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.* 27, 419–451.
- Guillery, R.W., and Sherman, S.M. (2002). Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. *Neuron* 33, 163–175.
- Lund, J.S., Hendrickson, A.E., Ogren, M.P., and Tobin, E.A. (1981). Anatomical organization of primate visual cortex area VII. *J. Comp. Neurol.* 202, 19–45.
- Miller, R. (2002). Wheels within wheels: circuits for integration of neural assemblies on small and large scales. In *Cortical Areas: Unity and Diversity*, R. Miller and A. Schüz eds. (London: Taylor and Francis), pp. 423–458.
- Rockland, K.S. (1997). Elements of cortical architecture: hierarchy revisited. In *Cerebral Cortex*, Vol. 12, Extrastriate Cortex in Primates, K.S. Rockland, J.H. Kaas, and A. Peters eds. (New York: Plenum Press), pp. 1243–1293.
- Shipp, S. (2003). The functional logic of cortico-pulvinar connections. *Phil. Trans. R. Soc. Lond. B* 358, 1605–1624.
- Shipp, S. (2005). The importance of being agranular: a comparative account of visual and motor cortex. *Phil. Trans. R. Soc. Lond. B* 360, 797–814.
- Sillito, A.M., Cudeiro, J., and Jones, H.E. (2006). Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci.* 29, 307–316.
- Thomson, A.M., and Bannister, A.P. (2003). Interlaminar connections in the neocortex. *Cereb. Cortex* 13, 5–14.
- West, D.C., Mercer, A., Kirchhecker, S., Morris, O.T., and Thomson, A.M. (2006). Layer 6 cortico-thalamic pyramidal cells preferentially innervate interneurons and generate facilitating EPSPs. *Cereb. Cortex* 16, 200–211.
- Yoshimura, Y., Dantzker, J. L. M., and Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature* 433, 868–873.

Department of Anatomy and Institute of Ophthalmology, University College London, Bath Street, London EC1V 9EL, UK. E-mail: s.shipp@ucl.ac.uk

Correspondences

Rapid advancement of spring in the High Arctic

Take T. Høye^{1,2}, Eric Post³, Hans Meltofte¹, Niels M. Schmidt^{1,4} and Mads C. Forchhammer^{1,4}

Despite uncertainties in the magnitude of expected global warming over the next century, one consistent feature of extant and projected changes is that Arctic environments are and will be exposed to the greatest warming [1]. Concomitant with such large abiotic changes, biological responses to warming at high northern latitudes are also expected to outpace those at lower latitudes. One of the clearest and most rapid signals of biological response to rising temperatures across an array of biomes has been shifts in species phenology [2–4], yet to date evidence for phenological responses to climate change has been presented from most biomes except the High Arctic [3]. Given the well-established consequences for population dynamics of shifts in the timing of life history events [5,6], it is essential that the High Arctic be represented in assessments of phenological response to climate change. Using the most comprehensive data set available from this region, we document extremely rapid climate-induced advancement of flowering, emergence and egg-laying in a wide array of species in a high-arctic ecosystem. The strong responses and the large variability within species and taxa illustrate how easily biological interactions may be disrupted by abiotic forcing, and how dramatic responses to climatic changes can be for arctic ecosystems.

Most long-term records of phenological events are from north-temperate environments. Recent comprehensive studies from this region have reported

campuses: the original Yoshida campus to the north and a newer southern campus at Uji. The division of Biological Sciences is located mainly at Yoshida, but the newer Institute for Chemical Research on the Uji campus includes molecular biology laboratories and recently became one of two centres for the

national GenomeNet service (www.genome.ad.jp).

What do the next 100 years hold for Kyoto University? Scientific policy is complex in Japan. Many of the buildings at Kyoto University are in dire need of repair, yet the government prefers to build smart

new universities in the countryside to attract people out of the cities as part of their decentralisation strategy. Although students still know that graduating from Kyoto University will help them towards their dream job, some younger researchers may be tempted to try their luck at one of the bright and shiny new places.

Primer

Synaptic plasticity Charles F. Stevens and Jane Sullivan

Most neurobiologists believe that memories are encoded in the pattern of cortical synaptic strengths, that is, in the effectiveness with which one neuron in the brain communicates with another at each of its synapses. As the human cortex contains about 10^{14} synapses (100 trillion), a lot of information could be stored.

Understanding how memory works is, by general agreement, one of neurobiology's central problems. Naturally, then, many neurobiologists are concerned with how synaptic strengths are regulated to store information; that is, they are interested in synaptic plasticity. Because long-term potentiation (LTP) and long-term depression (LTD) are the only mechanisms known for the persistent regulation of synaptic strength in neocortex, these two forms of synaptic plasticity have attracted great attention in recent years.

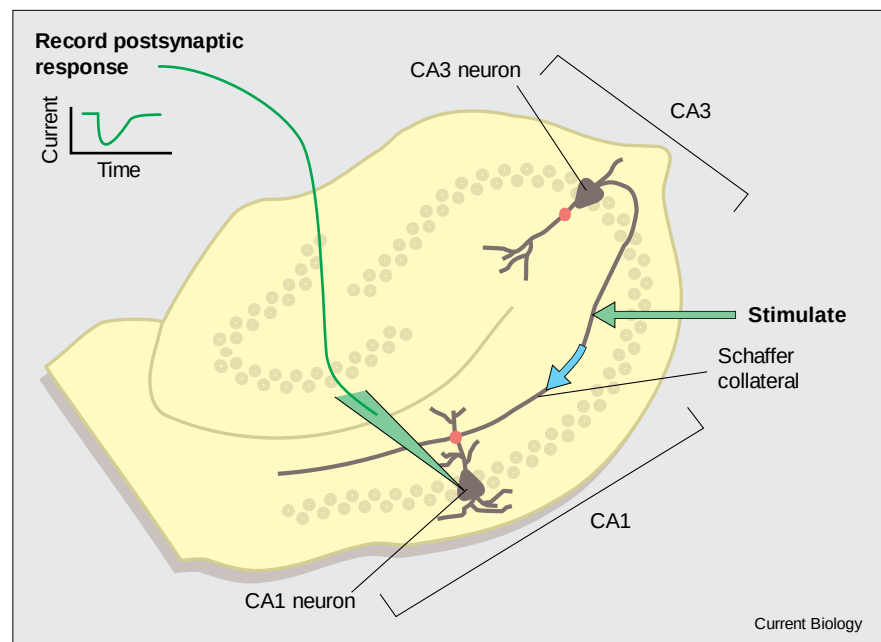
The hippocampus has been identified as a brain region concerned with encoding memories and is therefore typically, but not exclusively, used to study LTP/LTD. The standard preparation for studies of cellular mechanisms of plasticity is slices of hippocampus. By stimulating a particular pathway (the Schaffer collaterals comprised of axons from pyramidal cells in a region of

hippocampus known as 'CA3') and recording from any of the pyramidal cells in the hippocampal region designated 'CA1', the electrophysiologist can study direct (monosynaptic) excitatory synaptic connections (Figure 1).

Suppose we stimulate the Schaffer collateral fibers and record from a typical CA1 pyramidal cell. If the Schaffer collaterals are stimulated once every minute, we find that the size of the postsynaptic response recorded from the CA1 pyramidal cell remains constant from stimulus to stimulus (Figure 2). Now we use

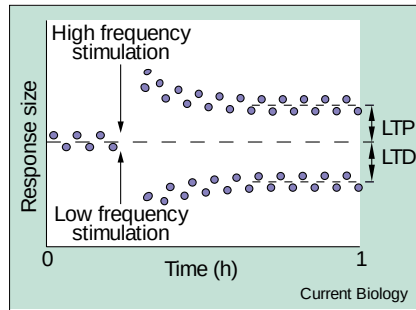
the synapses heavily: stimulate 100 times in a second. After this intense stimulation the size of the postsynaptic response is larger and remains so for the remainder of the hour-long experiment; this maintained increase in postsynaptic size is LTP. A different kind of synaptic use can, however, have the opposite effect. Instead of giving the strong 100-stimulus burst, had we stimulated once every second for 15 minutes, we would have found that the postsynaptic response got smaller and stayed that way for the rest of the experiment; this is LTD.

Figure 1



The hippocampal slice preparation has a very simple structure and serves as the synaptic physiologist's equivalent of the neuromuscular junction. Postsynaptic

responses can be recorded from any CA1 pyramidal neuron by stimulating the Schaffer collateral pathway, which is made up of axons from pyramidal neurons in the CA3 region.

Figure 2

The strength of a synapse (that is, the size of the postsynaptic response to presynaptic stimulation) remains constant if the synapse is used only once a minute. If, however, the synapse is stimulated 100 times a second for one second, the size of the response increases (LTP). If the synapse is stimulated once a second for 15 minutes, the size of the response decreases (LTD).

Our aim is to survey briefly what is known and, more importantly, what is not known about LTP and LTD. Because these phenomena are based on changes in synaptic strength we must, however, begin by reviewing the mechanisms of synaptic transmission at central synapses.

Synaptic transmission

As a single Schaffer collateral axon runs through the hippocampus it makes many synapses on dendritic spines of different CA1 pyramidal cells (Figure 3). The enlargement of the axon that constitutes the presynaptic element is called a bouton and it lies immediately adjacent to the head of a spine, which constitutes the postsynaptic element. The bouton membrane facing the spine head has a thickened area, called the active zone, to which about a dozen synaptic vesicles containing glutamate are docked; dozens of reserve synaptic vesicles wait in the bouton behind the active zone for their turn to dock. The spine head membrane directly opposite the active zone is also specialized (it appears dark in electron micrographs) to form the postsynaptic density. The extracellular space between the active zone and the

postsynaptic density is designated the synaptic cleft.

When a nerve impulse arrives at the bouton it causes a brief increase in the local calcium concentration. This calcium binds to special sites associated with the docking proteins and can cause an exocytotic event: a synaptic vesicle fuses with the membrane of the active zone and dumps its glutamate into the synaptic cleft. This fusion event is probabilistic and unlikely so that, more frequently than not, no exocytosis occurs when a nerve impulse arrives at the bouton. The chances that an exocytotic event will occur are roughly proportional to the number of vesicles that are docked to the active zone.

The glutamate released from a vesicle fusion event diffuses essentially instantaneously across the synaptic cleft and binds to glutamate receptors embedded in the postsynaptic membrane. The bound glutamate induces a conformational change in the receptors that are responsible for the excitatory postsynaptic current (EPSC) — the signal finally produced by the presynaptic nerve impulse.

Although molecular biologists have discovered an impressive array of glutamate receptor types, the receptors present at hippocampal synapses can be placed in two quite different categories: the AMPA and NMDA types. (AMPA and NMDA are abbreviations for the names of two glutamate analogs that pharmacologists used to classify receptor types before the gene families were known.) AMPA-type glutamate receptors (often called ‘AMPA receptors’ even though glutamate is the natural ligand) are responsible for normal synaptic transmission. When these AMPA receptors bind glutamate, they open an ion-selective pore and permit the influx of sodium ions that constitutes most of the EPSC.

The NMDA-type glutamate receptors have two important properties. First, they have a clever

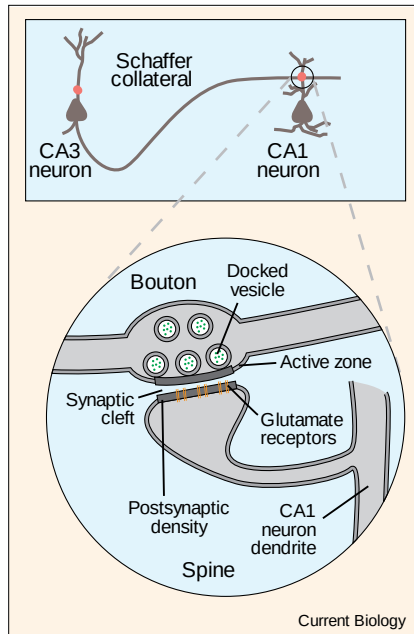
mechanism that permits them to sense simultaneously the presence of glutamate in the synaptic cleft and the postsynaptic voltage, so that they permit current to flow through their pore only when glutamate has been released presynaptically and other synapses have been activated (producing the required voltage change). In other words, NMDA receptors are coincidence detectors. Second, the NMDA receptor pore is very permeable to calcium ions, so that this receptor class can permit a significant calcium influx and a consequent increase in calcium concentration in the spine head.

LTP and LTD

Now we can start to explain what is known about LTP and LTD. Despite an enormous literature, hardly any feature of synaptic plasticity is generally accepted. Traditionally one distinguishes between mechanisms of induction (how is the plasticity triggered?), expression (what changes in a synapse to make it stronger or weaker?) and maintenance (what are the cellular and molecular mechanisms for keeping synaptic strength at the required level?).

The trigger for both LTP and LTD is calcium influx through NMDA receptor pores. This is one of the few universally accepted properties of LTP/LTD. The coincidence-detection property of the NMDA receptors is used here, so that LTP requires the simultaneous activation of many synapses. What determines if calcium influx through NMDA receptor pores results in LTP or LTD? Most workers believe that the most important thing is the quantity of calcium that enters: a little calcium gives LTD and a lot of calcium produces LTP. But the history of synaptic use, in addition to the calcium concentration, also seems to play a part in deciding what form of plasticity occurs. The big question ‘What are the exact requirements for the induction of LTP and LTD?’ has not been answered.

Figure 3



The CA3–CA1 synapse is made up of a presynaptic bouton and a postsynaptic density. Vesicles filled with the neurotransmitter glutamate (green) are docked to the active zone of the bouton. Glutamate receptors (orange) are embedded within the membrane of the postsynaptic density.

Calcium influx is the trigger for synaptic plasticity. What are the calcium sensors? The postsynaptic density has associated with it a variety of calcium and calmodulin-dependent kinases and phosphatases as well as other calcium-regulated enzymes. Many of these have been proposed to provide the next link in the chain of signaling events, but no clear story has emerged. The big question ‘What are the signaling pathways downstream from increases in calcium concentration?’ has not been answered.

As for mechanisms of expression, how can synapses get stronger in LTP and weaker in LTD? A synapse can get stronger: by increasing release probability; by increasing the quantity of glutamate in each vesicle; by increasing the number of glutamate receptors in the postsynaptic membrane; or by increasing the size of the response produced by each glutamate receptor. LTD could be the

reverse of any of these. The quantity of glutamate per vesicle is believed (without real evidence) to be fixed, but all of the other possibilities have their advocates. Notice that increasing release probability is a presynaptic mechanism, whereas increasing the response to the glutamate released (the last two of the possibilities mentioned above) is a postsynaptic mechanism.

The current heated debate on the nature of expression focuses not so much on the detailed mechanisms as on the locus. Much evidence can be adduced for both presynaptic and postsynaptic loci for plasticity expression, but no consensus has yet developed. The problem with almost all of the experiments on this question of expression is that they are indirect and interpretation always depends on assumptions that are not directly verified. The positions of individual scientists on these issues come down to the subjective estimate of which of the unverified assumptions are most likely to be correct.

The idea that has probably attracted the most attention in the last several years is the ‘silent synapse hypothesis’. According to this idea, many synapses have only NMDA receptors and LTP occurs by the addition of AMPA receptors; LTD, according to this view, could be the opposite. The big question ‘Is LTP/LTD expressed presynaptically or postsynaptically and what are the cell and molecular biological mechanisms responsible?’ has not been answered.

As for the maintenance of synaptic strength, something has to change at synapses to keep them at the right strength. Most of those who work on synapses believe (without direct evidence) that the ultimate change is structural — cortical synapses grow or shrink to stay stronger or weaker — but this presumably takes some time and both LTP and LTD occur in a matter of minutes. Protein kinase A has recently become a popular candidate for a role in the longer term maintenance of synaptic strength.

The big question ‘What are the molecular mechanisms responsible for maintaining synaptic strength at the various time scales?’ has not been answered.

One of the key properties claimed for LTP/LTD is that the changes in strength are not generalized but are specific to the synapses that have been activated. The degree of specificity has been questioned in recent years and spread of LTP/LTD has been demonstrated (some would say claimed) to spread both presynaptically (to neighboring synapses made by the same axon) and postsynaptically (to neighboring synapses made on the same dendrite by different axons). The big question ‘To what extent is LTP/LTD synapse-specific and what are the mechanisms that regulate the extent of this specificity?’ has not been answered.

Memory

LTP and LTD are the only really long-lasting changes in synaptic strength (outside of those that occur during development) known in the neocortex. Are these phenomena responsible for learning and memory? Although most learning and memory experts (about 60% according to our latest polls) believe LTP and LTD are the basis for certain forms of memory, this is the really big unanswered question.

Key references

- Abraham WC, Bear MF: **Metaplasticity: the plasticity of synaptic plasticity.** *Trends Neurosci* 1996, **19**:126-130.
- Engert F, Bonhoeffer T: **Synapse specificity of long-term potentiation breaks down at short distances.** *Nature* 1997, **388**:279-284.
- Goda Y: **A common cascade for long-term memory.** *Curr Biol* 1995, **5**:136-138.
- Nicoll RA, Malenka RC: **Contrasting properties of two forms of long-term potentiation in the hippocampus.** *Nature* 1995, **377**:115-118.
- Stevens CF, Wang Y: **Facilitation and depression at single central synapses.** *Neuron* 1995, **14**:795-802.

Address: Molecular Neurobiology Laboratory and Howard Hughes Medical Institute, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA.

Primer

Visual psychophysics Oliver Braddick

Visual processes in the brain enable us to make sense of the complex, ever-changing array of shapes and colours that arrives at our eyes. Psychophysics is the research strategy that seeks to understand these processes, by testing the relationship between the 'psychic' — what an observer sees and reports — and variations in the physical pattern of light and colour.

For 19th century pioneers such as Weber, Helmholtz and Mach, psychophysics provided the only window on how sensory information was transmitted and organized in the nervous system. In the mid-20th century, however, it became possible to measure neural responses directly — as single-cell responses in animals and, more recently, by functional imaging of local activation in the human brain. But these advances do not make the psychophysical approach obsolete; rather, they require psychophysics as a partner. In looking at the responses of individual neurons alone, it can be hard to understand what part they play in overall visual function. Psychophysical experiments are needed to understand how thousands of neurons act together to register shape, a coloured area, or a pattern of movement — or indeed, to integrate these into a whole image, such as when we perceive a running, brown horse. In addition, one of the deepest questions facing neuroscience is how neural activity is related to our experience of seeing; by definition, this requires the parallel study of neurophysiology and psychophysics.

Simplifying the visual process

Our visual experience is rich and multidimensional. If we simply ask observers to report what they see, it

is hard to discipline the results into any systematic and reproducible form that can test hypotheses about how neural signals carry different kinds of visual information. So, to achieve this, psychophysical experiments usually restrict the observer's report to some very cut-and-dried judgment, such as: 'Did the pattern of dots appear to move up or down?'. Such a 'forced-choice' judgment, repeated many times with a selection of different visual stimuli, is often analysed and presented in graph form as a 'psychometric function' (see yellow box).

Psychometric functions

A psychometric function is a graph plotting the proportion of an observer's judgments in one of two categories (for example, 'moved up') against the variation in the physical stimulus (for example, how far the pattern of dots was displaced). Large negative (downward) displacements are never judged to be 'up', and large positive displacements give 100% 'up' judgments. In between these extremes, judgments are not perfectly reliable. A particular displacement of the dots will be judged as 'up' on, say, only 65% of occasions. This gives the characteristic S-shaped form of the graph (see Fig.1).

The function yields at least two distinct kinds of information. Its slope reveals the sensitivity of the neural mechanism underlying the judgment: a steep function means that the system can respond to a small stimulus variation, producing a reliable change in the judgement. In this example, sensitivity to small displacements would reflect the behaviour of motion-detecting neurons in the human visual system. Sensitivity is often expressed by taking a level of observers' judgments, such as 75%, and defining the stimulus needed to produce this level as the 'threshold' for discrimination (in this case, a displacement threshold). But such a 'threshold' is, in fact, an arbitrary point on a continuous function; it does not imply a level at which the discrimination suddenly becomes possible.

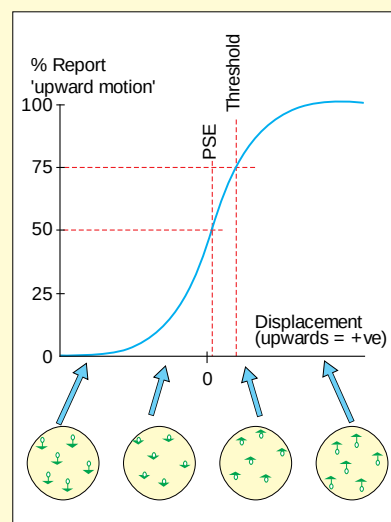
A second kind of information is given by the 50% point, at which 'up' and 'down' judgments are equally probable. This 'point of subjective equality' (PSE) may sometimes deviate from the point where the physical stimulus is exactly balanced (here, not moving). Such a bias in the

The aim of forced-choice experiments is to give an objective measure of the information from the visual stimulus that is being registered by the sensory system, unaffected by any expectations or biases that the observer may have. In particular, two-interval forced-choice experiments (see pink box) make it possible to study both simple and complex visual discriminations in a standardized, objective way. Rather than having to describe some aspect of what they see, observers simply have to choose between two slightly different versions of the same visual stimulus. Any difference that can be

observer's perceptual system can occur, for instance, when one set of motion-sensitive neurons has been adapted by prolonged exposure to a particular direction.

Motion is just one stimulus property to which visual neurons give specific responses. Psychophysical experiments can also manipulate colour, orientation, or spatial frequency (fine detail *versus* large-scale form) to explore the properties of neurons that encode these other aspects of the visual array.

Figure 1



Psychometric function for discrimination between two opposed stimuli (in this case, opposite directions of motion).

How do we see the whole picture?

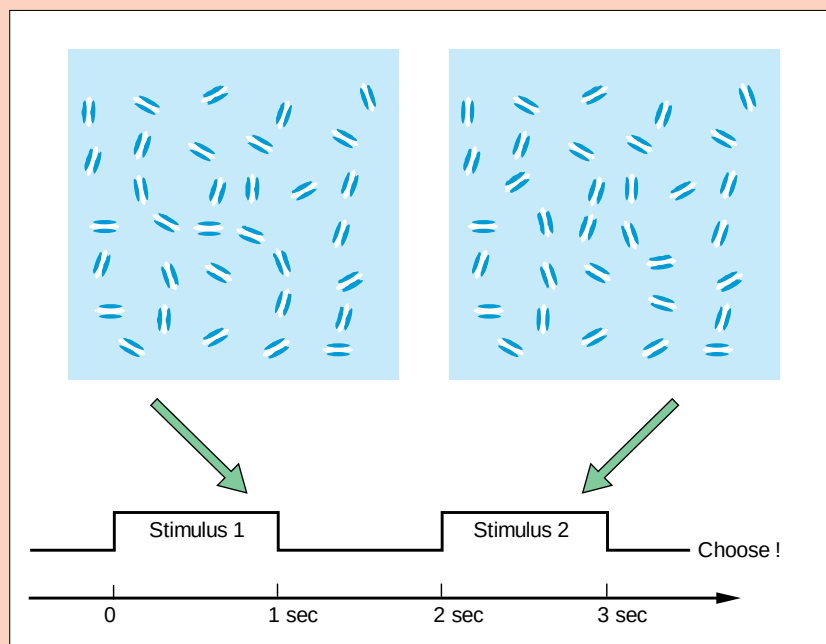
Gestalt psychologists long ago described how visual elements that form a continuing chain are perceived as a group. But how can the complex, subjective sense of grouping be studied as a concrete phenomenon? After all, people often see pattern even in random arrangements, like faces seen in the fire.

To answer this question, David Field and his colleagues exploited the psychophysical method of 'two-interval forced-choice'. On each trial, or 'interval', two patterns were briefly displayed, one after the other, 1 second apart. The observer had to report which of the patterns contained a 'path' of linked elements (see Fig.2). Observers might have their individual biases, but in this

experiment they have to choose which of the two gives a stronger sense of 'path'. As the path occurs in each interval equally often, any bias towards choosing 'first' or 'second' is unrelated to the visual property under study. Making 50% correct judgments is 'chance' performance, indicating that the observer is unable to detect the path.

Field and colleagues used such data to plot psychometric functions of 'percent correctly reported grouped elements' against 'the angle between neighbouring elements in the path', and inferred the pattern of connections by which angle-sensitive neurons may activate their neighbours, to enhance the perception of the elements as being grouped.

Figure 2



A two-interval forced-choice sequence as used in the experiment of Field and colleagues on grouping neighbouring elements. In the sequence shown, the correct response would be that the 'path' (an S-shaped path slightly left of centre) is present in the first interval.

detected between the two stimuli — what experimental psychologists call a 'cue' — is fair game for the observers in their attempt to make the required choice. The skill of experimental design in psychophysics lies in constructing pairs of stimuli where the only systematic difference between them — the cue — is that which should be detected by the mechanism the

experimenter wants to study; so, there is no opportunity for the observers to 'cheat', either consciously or unconsciously.

Learning to recognize visual cues

When the experimenter has successfully constructed a task that depends on a specific visual cue, the observers have to recognize that cue and harness their decisions to it. Any

stimulus must activate many millions of neurons in the visual areas of the observer's brain. All recordings from neurons show that their responses are quite 'noisy': the signals vary randomly even on occasions when exactly the same stimulus is presented. The observer, somehow, has to pick up which part of this torrent of fluctuating signals reflects the cue that is the focus of the experiment. It is hardly surprising that observers have to learn what they are supposed to be doing.

Psychophysical experimenters have always begun with blocks of practice trials, intended to ensure that observers understand the point of the task and settle down to a stable level of performance. But many recent experiments have shown that observers can learn to make a particular discrimination increasingly well over periods of many days. This learning can be extraordinarily specific. For example, practice that leads to a striking improvement in finding oriented line segments in the upper right visual field may leave unchanged performance on the same task in lower right or upper left regions. 'Practised' observers must be learning to couple their judgments to patterns of activity occurring in some quite specific group of neurons, early in the chain of brain areas that process visual information. In other words, if the aim of a psychophysical experiment is to test what organized pathways are present in the brain, it may be more realistic to consider it as testing 'what organized pathways can the brain create, given time, to meet the demands of this experiment?'

Psychophysics is beginning to come to grips with the flexibility of the visual system, flexibility that we exploit when we learn to tell a painting by Monet from one by Renoir, or a perfect blackberry from one that is not quite ripe.

The task for the psychophysical observer, then, is defined not just by instruction but by training on the job. This complicates the interpretation

of experiments, but it also opens up a new range of possibilities. Verbal instructions can be delivered only to human beings, but many other species can be trained in certain behaviours. For example, forced-choice experiments can be set up in which pigeons define their thresholds for light perception by pecking at one of two illuminated keys; kittens can show texture discriminations by jumping onto patterned platforms; and macaque monkeys reveal by lever-pulling or eye movements that they have colour, contrast, and motion thresholds similar to those of humans. In each case, the correct response has to be reinforced by food or some other reward, but once the association has been learned, animal 'observers' can produce psychometric functions over hundreds of trials, just like their human counterparts.

Nerve impulses and perception

The value of 'animal psychophysics' is not simply what we can discover

about the perceptual world of species other than our own. It also provides the link to directly measured neural function, for these other species are ones in which we can probe the living brain with microelectrodes. The most exciting possibilities come about when the activity of single brain cells is recorded at the same time as the trained animal makes its psychophysical judgments.

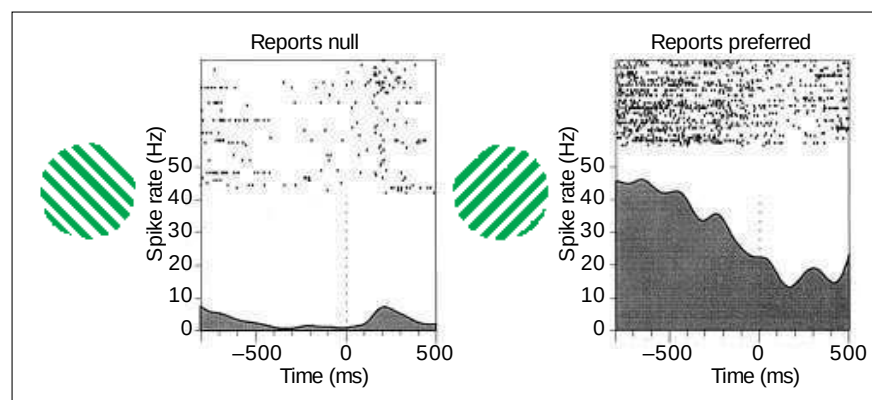
Psychometric functions show how what we think we've seen depends on the physical stimulus. But our sensations depend not only on effects from the physical world outside, but also on processes within ourselves. One example is binocular rivalry. Normally, the two eyes work together to integrate two views of the same scene, but if the two eyes see completely incompatible views (for example, stripes at right angles to each other) they compete in perception; human observers report seeing one or other pattern dominate. The dominant pattern

switches from time to time, but a blend of the two is never seen. Nikos Logothetis and his colleagues have shown that a monkey, trained to press different levers for two stripe patterns, alternates the two with a pattern just like that from a human observer reporting binocular rivalry.

Microelectrodes placed in visual brain areas during this procedure can locate cells which respond vigorously when the monkey is 'reporting' one stripe pattern as dominant, but are silent when the monkey is pressing the other lever, even though the physical patterns presented to the eyes are unchanged (see Fig. 3). In other words, the pattern of firing of these neurons corresponds to what the monkey 'tells' the experimenter it has seen. Intriguingly, some cells in brain areas V1, V2, V4 and V5 show this variation, whereas others respond in a way that is apparently not correlated with the perceived dominance of the stimuli.

This line of experimentation suggests that perceptual awareness reflects the activity of some neurons more than others, even at the same level of the visual pathway. Such experiments, in which the methods of psychophysics and neurophysiology work together, offer the promise of mapping the presently mysterious links between brain and perception.

Figure 3



Responses of a neuron in brain area V1/V2, associated with perceived dominance in perception of a particular orientation of stripes in binocular rivalry. The results were recorded by Leopold and Logothetis from a macaque monkey, while it pressed a key to report seeing a certain orientation as dominant. The neuron responded strongly to stripes at an angle of 45° presented to either eye ('preferred orientation'), and not at all to stripes at 135° ('null orientation'). The graphs show responses when the animal viewed 45° stripes through one eye and 135° stripes simultaneously through the other. The graph

on the left shows the low rate of responses occurring during the 800 ms before, and 500 ms after the key press, indicating that the animal was seeing the null orientation; the graph on the right shows the much higher activity around the time of a key press, indicating the preferred orientation. In each graph, the vertical ticks indicate individual action potentials (each row being a separate time interval containing a key press); the curve below is the averaged rate of action potentials over many intervals. (Reproduced with permission from *Nature*.)

Key references

- Watt R J: **Psychophysics and visual descriptions; Psychophysics and visual interpretation.** In *Understanding Vision*. London: Academic Press; 1991:245–292.
- Field DJ, Hayes A, Hess RF: **Contour integration by the human visual system: evidence for a local 'association field'.** *Vision Research* 1993, **33**:173–193.
- Karni A, Sagi, D: **Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity.** *Proc Natl Acad Sci USA* 1991, **88**:4966–4970.
- Leopold DA, Logothetis NK: **Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry.** *Nature* 1996, **379**:549–554.

Address: Department of Psychology, University College London, Gower St, London WC1E 6BT, UK.