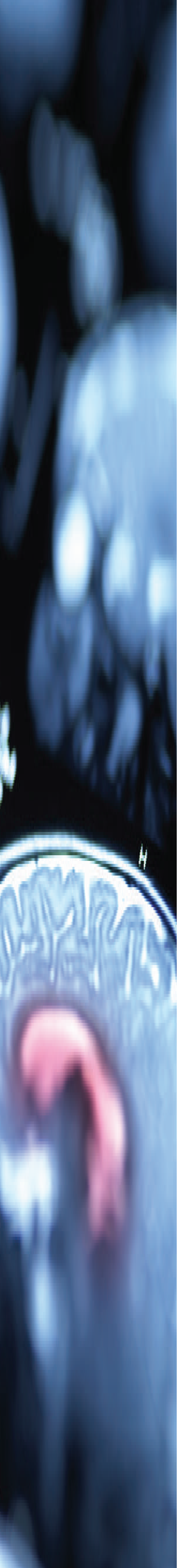


Generalization in vision and motor control

Tomaso Poggio[1] & Emilio Bizzi[1,2]

[1] *McGovern Institute, Department of Brain and Cognitive Sciences, Center for Biological and Computational Learning, Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA (e-mail: tp@ai.mit.edu)*
[2] *European Brain Research Institute, Via del Fosso di Fiorano, Roma 00143, Italy (e-mail: ebizzi@mit.edu)*

Learning is more than memory. It is not simply the building of a look-up table of labelled images, or a phone-directory-like list of motor acts and the corresponding sequences of muscle activation. Central to learning and intelligence is the ability to predict, that is, to generalize to new situations, beyond the memory of specific examples. The key to generalization, in turn, is the architecture of the system, more than the rules of synaptic plasticity. We propose a specific architecture for generalization for both the motor and the visual systems, and argue for a canonical microcircuit underlying visual and motor learning.

Arguably, the problem of learning represents a gateway to understanding intelligence in brains and machines, to discovering how the human brain works and to making intelligent machines that learn from experience. What distinguishes nontrivial learning from memory is the ability to generalize: that is, to apply what has been learned from limited experience to new situations. Memory bears the same relationship to learning as a dry list of experimental measurements does to a predictive scientific theory. The key question addressed here — from the perspective of the visual and motor systems — is what are the brain mechanisms for such generalization?

Imagine looking for the phone in a new hotel room. Your visual system can easily spot it, even if you have never seen that particular phone or room before. So, learning to recognize is much more than straightforward pixel-by-pixel template matching. Visual recognition is a difficult computational problem, and it is a key problem for neuroscience. The main computational difficulty is that the visual system needs to generalize across huge variations in the appearance of an object; for instance, owing to viewpoint, illumination or occlusions. At the same time, the system needs to maintain specificity; for example, to identify a particular face among many similar ones.

A similar ability to generalize is key to motor learning. Consider practicing how to hit a tennis ball: having learned to play a specific shot, you must then be able to use it under new conditions, adapting to changes in the spin on the incoming ball, the speed and direction of your opponent's shots, the position of your body with respect to the ball, and so on. No two shots can be exactly the same, requiring a generalization ability of our motor program that can involve the modulation of thousands of motor units in new, adaptive ways.

In abstract terms, generalization is the task of synthesizing a function that best represents the relationship between an input, x, and an output, y — an image and its label, say, or a desired equilibrium position of an arm and the set of forces necessary for attaining it — by learning from a set of 'examples', x_i, y_i. In this formulation, the problem of learning is similar to the problem of fitting a multivariate function to a certain number of measurement data. The key point is that the function must generalize. Generalization in this case is equivalent to the ability of estimating correctly the value of the function at points in the input space at which data are not available — that is, of interpolating 'correctly' between the data points. In a similar way, fitting experimental data can, in principle, uncover the underlying physical law, which can then be used in a predictive way. In this sense, the process of learning distils predictive 'theories' from data; that is, from experience.

The modern mathematics of learning[1] gives a precise definition of generalization and provides general conditions that guarantee it. It also implies that the ability to generalize in the brain depends mostly on the architecture of the networks used in the process of learning, rather than on the specific rules of synaptic plasticity. (The latter are reviewed in this issue by Abbott and Regehr, page 796.)

Here, we highlight a network architecture supporting the ability to generalize in the visual and motor systems. Neurons at various levels of the visual cortex are generally tuned simultaneously to multiple attributes; that is, they respond to a particular pattern of their inputs, and the frequency of the firing follows a 'tuning curve', with a maximum for specific values of each of the attributes (together representing an optimum stimulus for the neuron), such as a particular direction of movement and a specific colour and orientation (Fig. 1 shows tuning for specific object views, each characterized by many parameters; see review in this issue by Tsodyks and Gilbert, page 775). We describe how a linear combination of the activities of such neurons can allow generalization, on the condition that the tuning is not too sharp, and that the weights of such a linear combination are what changes during learning. We then turn to the motor system and show that a linear combination of neural modules — each module involving several motor neurons innervating a coherent subset of muscles which together generate a force field — is mathematically equivalent to the linear combination of tuned neurons described for the visual system. Finally, we propose that the necessarily broad tuning of motor and visual neurons might be based on a canonical microcircuit repeated throughout different areas of cortex.

Generalization mechanisms in the visual system

Older mental models of how vision might work used the simple notion of 'computation through memory'. The classic example is the 'grandmother' theory for vision, in which visual recognition relies on 'grandmother' neurons responding selectively to the precise combination of visual features that are associated with one's grandmother. This theory was not restricted to vision: the same basic idea surfaced for other sensory modalities, for example in motor control, where it is called 'motor tapes'. These ideas were attractive because of their simplicity: they replace complex information processing with the simpler task of accessing a memory.

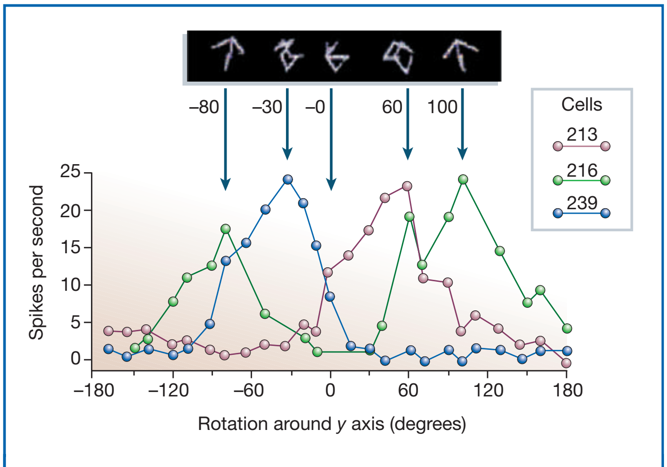


Figure 1 Tuned units in inferotemporal cortex. A monkey was trained to recognize a three-dimensional 'paperclip' from all viewpoints (pictured at top). The graph shows tuning to the multiple parameters characterizing each view summarized in terms of spike rate versus rotation angle of three neurons in anterior inferotemporal cortex that are view-tuned for the specific paperclip. (The unit corresponding to the green tuning curve has two peaks — to a view of the object and its mirror view.) A combination of such view-tuned neurons (Fig. 2) can provide view-invariant, object specific tuning as found in a small fraction of the recorded neurons. Adapted from Logothetis *et al.*[13].

The basic problem with these models is, of course, generalization: a look-up table cannot deal with new events, such as viewing a face from the side rather than the front, and it cannot learn in the predictive sense described earlier. One of the simplest and most powerful types of algorithm developed within learning theory corresponds to networks that combine the activities of 'units', each broadly tuned to one of the examples (Box 1). Theory (see references in Box 1) shows that a combination of broadly tuned neurons — those that respond to a variety of stimuli, although at sub-maximal firing rates — might generalize well by interpolating among the examples.

In visual cortex, neurons with a bell-shaped tuning are common. Circuits in infratemporal cortex and prefrontal cortex, which combine activities of neurons in infratemporal cortex tuned to different objects (and object parts) with weights learned from experience, may underlie several recognition tasks, including identification and categorization. Computer models have shown the plausibility of this scheme for visual recognition and its quantitative consistency with many data from physiology and psychophysics[2–5].

Figure 2 sketches one such quantitative model, and summarizes a set of basic facts about cortical mechanisms of recognition established over the last decade by several physiological studies of cortex[6–8]. Object recognition in cortex is thought to be mediated by the ventral visual pathway running from primary visual cortex, V1, over extrastriate visual areas V2 and V4 to the inferotemporal cortex. Starting from simple cells in V1, with small receptive fields that respond preferably to oriented bars, neurons along the ventral stream show an increase in receptive field size as well as in the complexity of their preferred stimuli. At the top of the ventral stream, in the anterior inferotemporal cortex, neurons respond optimally to complex stimuli such as faces and other objects. The tuning of the neurons in anterior inferotemporal cortex probably depends on visual experience[9–19]. In addition, some neurons show specificity for a certain object view or lighting condition[13,18,20–22]. For example, Logothetis *et al.*[13] trained monkeys to perform an object recognition task with isolated views of novel three-dimensional objects ('paperclips'; Fig. 1). When recording from the animals' inferotemporal cortex, they found that the great majority of neurons selectively tuned to the training objects were view-tuned (see Fig. 1) to one of the training objects. About one tenth of the tuned neurons were view-invariant, consistent with an earlier computational hypothesis[23].

In summary, the accumulated evidence points to a visual recognition system in which: (1) the tuning of infratemporal cortex cells is obtained through a hierarchy of cortical stages that successively combines responses from neurons tuned to simpler features; and (2) the basic ability to generalize depends on the combination of cells tuned by visual experience. Notice that in the model of Fig. 2, the tuning of the units depends on learning, probably unsupervised (for which several models have been suggested[24]; see also review in this issue by Abbott and Regehr, page 796), since it depends only on passive experience of the visual inputs. However, the weights of the combination (see Fig. 3) depend on learning the task and require at least some feedback (see Box 2).

Thus, generalization in the brain can emerge from the linear combination of neurons tuned to an optimal stimulus — effectively defined by multiple dimensions[25,23,26]. This is a powerful extension of the older computation-through-memory models of vision and motor control. The question now is whether the available evidence supports the existence of a similar architecture underlying generalization in domains other than vision.

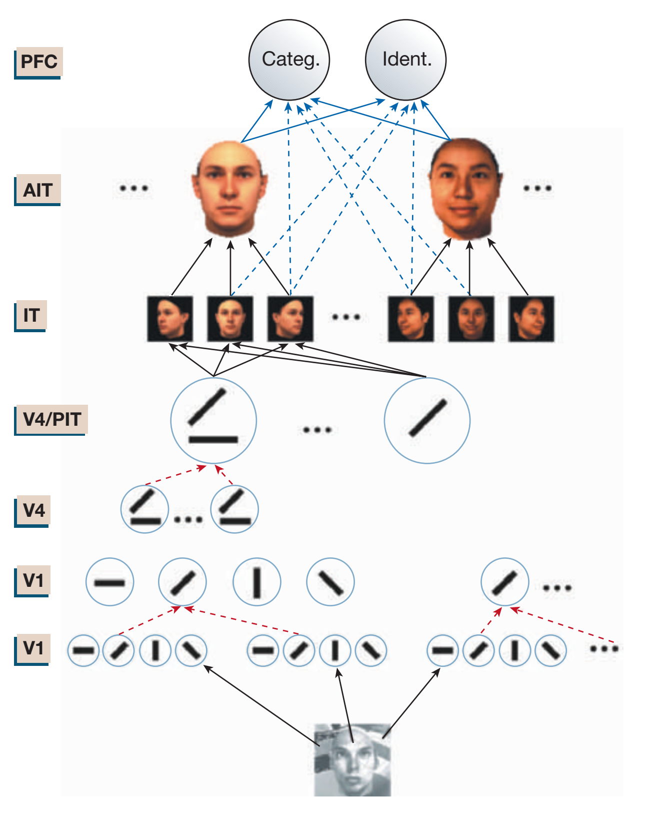


Figure 2 A model of visual learning. The model summarizes in quantitative terms other models and many data about visual recognition in the ventral stream pathway in cortex. The correspondence between the layers in the model and visual areas is an oversimplification. Circles represent neurons and arrows represent connections between them; the dots signify other neurons of the same type. Stages of neurons with bell-shaped tuning (with black arrow inputs), that provide example-based learning and generalization, are interleaved with stages that perform a max-like operation[3] (denoted by red dashed arrows), which provides invariance to position and scale. An experimental example of the tuning postulated for the cells in the layer labelled inferotemporal in the model is shown in Fig. 1. The model accounts well for the quantitative data measured in view-tuned inferotemporal cortex cells[10] (J. Pauls, personal communication) and for other experiments[55]. Superposition of gaussian-like units provides generalization to three-dimensional rotations and together with the soft-max stages some invariance to scale and position. IT, infratemporal cortex, AIT, anterior IT; PIT, posterior IT; PFC, prefrontal cortex. Adapted from M. Riesenhuber, personal communication.

Box 1

Learning and generalization with tuned, gaussian-like units

Basis functions

The problem of learning from examples can be formulated as a problem of function approximation with the property of generalization (robustness to noise is a special case of the ability to generalize). A classical and simple mathematical approach to solving it is regularization: the function *f* learned from the data minimizes the error on the training set subject to certain 'smoothness' constraints. An intriguing result is that the solution of the minimization problem above can be expressed as a linear combination of basis functions *k* centred on the examples $\bar{x}_i$ and depending on the new input vector $\bar{x}$:

$$f(\bar{x}) = \sum_{i}^{n} w_i k(\bar{x},\bar{x}_i),$$

where the $\bar{x}_i$, are the *n* (vector) examples and w_i are parameters to be determined (for example, learned) from the *n* example pairs $\bar{x}_i$, y_i, where $\bar{x}_i$ is the 'input' part of each example and y_i is its associated label or 'output'. The basis functions *k* are fixed functions, such as the gaussian function, of the input. Note that the centres $\bar{x}_i$ (the optimal stimuli for each of the basis functions) are simply learned from 'passive' visual experience without the need of feedback, whereas the weights w_i also depend on the y_i (corresponding to the feedback) and can also be learned by simple learning rules such as the delta rule or the covariance rule[60]. When the basis functions are radial gaussian functions, the network consists of units each tuned to one of the examples with a bell-shaped activation curve. Each 'unit' computes the distance $\bar{x} - \bar{x}_i$ of the input vector $\bar{x}$ from its centre $\bar{x}_i$ (that is, the dissimilarity of the input and the example stored in that unit) and then applies the function *k* to the dissimilarity value, that is, it computes the function $k(\bar{x} - \bar{x}_i)$. Notice that in the limiting case of $k(\bar{x} - \bar{y})$ being a delta function (for example, a very narrow gaussian function), the network becomes a look-up table, in which a unit gives a non-zero signal only if the input $\bar{x}$ exactly matches its centre $\bar{x}_i$: the network cannot generalize and becomes a simple memory. The equation above can always be rewritten as a feedforward network (Fig. 3a) with one hidden layer containing as many units as examples in the training set. The units of the hidden layer correspond to the basis functions and can be regarded as processors doing a specific operation; the parameters w_i correspond to the weight of the synapses from the units *k* to the output. The scalar output case above can be generalized to the multi-output case (for example, the approximation of vector fields), which is the general case and the relevant one for motor control (Fig. 3b). The function $f(\bar{x})$ is thus the superposition of local, tuned receptive fields such as gaussians; it is predictive; and it is also a smooth (the exact definition of 'smooth' depends on *k*) solution as it minimizes a smoothness constraint such as jerk[61], while being close to the examples. There are alternative ways to implement similar solutions within recurrent networks[25]. It is well known that regularization networks can be interpreted in bayesian terms[25,62] but detailed models of how general bayesian networks and graphical models may be implemented in the visual or motor system are lacking so far.

Time

We have described time-independent aspects of visual recognition and motor control, corresponding respectively to recognition of static images and to control of postures (say of an arm). In most of real life, recognition and motor control happen in time: we recognize actions and we control dynamic movements. The equations above, describing the superposition of 'prototypical' images or prototypical force fields, can be extended in time. Possibly the simplest such extension is provided by:

$$\bar{f}(x,t) = Y\bar{b}(x) = \sum_{i}^{n} b_i(x) g_i(t) \bar{y}_i$$

where $\bar{f}$ and $\bar{y}_i$ are vector fields and $g_i(t)$ the associated time dependence. This representation seems to be consistent with experimental data in motor control[54].

A similar description summarizes the model of Giese and Poggio[55] for the recognition of actions from image sequences. Sequence selectivity results from asymmetric lateral connections between the snapshot neurons in the form pathway (and between the optic flow pattern neurons in the motion pathway). With this circuitry, active snapshot neurons pre-excite neurons that encode temporally subsequent configurations and inhibit neurons that encode other configurations. Significant activity can arise only when the individual snapshot neurons are activated in the 'correct' temporal order. Simulations show that in the model, appropriate lateral connections for the 'correct' sequences can be learned robustly with a simple time-dependent hebbian learning rule from a small number of stimulus repetitions, consistent with psychophysical data[55].

Generalization mechanisms in the motor system

The architecture for generalization outlined for the visual system (Fig. 3a) leads to a stage of broadly tuned units. For any specific visual recognition task, there are many inputs (such as the photoreceptors) and just one output signal. In the computational architecture of the motor system, however, the flow of information is the opposite, with few inputs (discrete cortical commands from the fronto-parietal cortex) and many outputs (the interneurons and motorneurons in the spinal cord). For such architectures, the combination (with fixed weights set by learning) of neurons tuned by learning to optimal stimuli (with an activity dependent on the similarity between the input and the optimal stimulus) can be formally viewed (see legend of Fig. 3) as a combination (with weights depending on the input signal) of neural circuits or modules, each generating a (fixed) motor 'field' of muscle forces. The non-trivial equivalence may lead to novel experiments. It also suggests that the evidence in the literature about tuned neurons may be fully compatible with the apparently different reports supporting the combination of modules and associated force fields.

In the fronto-parietal cortical areas, arm-related, broadly directionally tuned neurons were first described by Georgopoulos *et al.*[27]. These neurons are related to arm movements and their tuning means that their frequency of discharge varies in an orderly fashion with the direction of movement. For each neuron, the discharge was most intense in a preferred direction resulting in a directional bell-shaped tuning curve. In the motor areas of the frontal lobe, neurons with similar preferred direction are interleaved with mini-columns having nearly orthogonal preferred directions[28]. This recent discovery indicates that the motor cortex is endowed with functional modular structures not unlike those described for the visual cortex[6,7], the somato-sensory cortex[8] and the auditory cortex[29]. Neuronal activity in the frontal cortical areas, such as the primary motor cortex, the supplementary motor areas and the dorsal premotor areas, change during adaptation and visuo-motor learning[30,31], and during exposure to mechanical loads[32–34]. In addition, during motor learning a significant number of cortical cells change their directional tuning.

While the significance of the information conveyed by the activity of broadly tuned cortical neurons remains hotly debated, here we put forward the hypothesis that the descending cortico-spinal impulses may represent signals (such as the components of the vector $b(\bar{x})$ in Fig. 3) that specify the activation for the modules in the spinal cord of vertebrates. Several kinds of modular spinal systems, consisting of circuits of interneurons, have been described. These range from central pattern generators and unit burst generators[35–37] to spinal motor primitives generating specific force fields and muscle synergies[38,39].

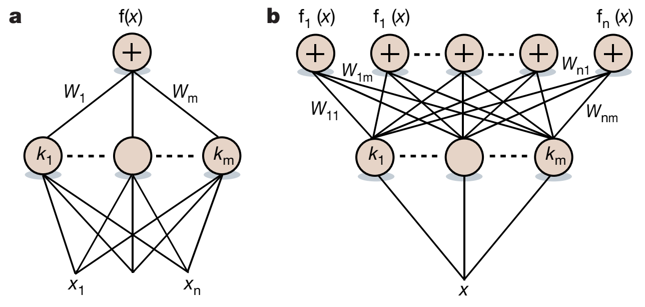


Figure 3 The generalization architectures of the visual and motor systems. **a**, In the case of vision, the single output signal is a combination of tuned unit activities $f(\bar{x}) = \sum_i^n w_i k_i(\bar{x})$. **b**, In the case of motor control, the output vector can be similarly written as $\bar{f}(\bar{x}) = \sum_i^n \overline{w}_i k_i(\bar{x})$ where each component of the output field is a combination of tuned unit activities. Here, $\overline{w}_i$ is the *vector* $w_{1i}, \ldots, w_{ni}$ of the weights associated with the tuned unit *i*. The same equation can also be read as a combination of the fields $\overline{w}_i$ with coefficients $k_i(\bar{x})$; that is, a combination of fields modulated by the activities of the tuned units. Thus, a combination of tuned units is formally equivalent to a combination of fields. The general description of the networks shown is given, rewriting the last equation, by $f(\vec{x}) = Wk(\vec{x})$, where W is a matrix and $k(\vec{x})$ is a vector with the tuned units as components. Notice that the factorization in terms of coefficients and basis function is not unique (when only the input and the outputs of the network are observed) since $Wk(\vec{x}) = Cb(\bar{x})$ where L is any matrix satisfying $W = CL$ and $b(\vec{x}) = L\bar{k}(\bar{x})$. An additional constraint (such as specifying which parameters of the network change with learning) is needed to fix L. The arbitrariness in the decomposition might explain apparent differences in the interpretations of some experiments. For instance, Thoroughman and Shadmer[50] conclude from behavioural data that the basis functions are gaussian and tuned to desired velocities, whereas cortical cells would presumably show a linear tuning as a function of velocity[27].

Box 2

Supervised and semi-supervised online learning

The distinction between supervised and unsupervised learning in biology can be tricky: there is a whole spectrum between the two. The learning process by which the neurons in the model of Box 1 get tuned more and more specifically to a particular value of a given attribute is unsupervised: it relies only on the inputs, and it does not require feedback about the correctness or incorrectness of the output. Several mechanisms have been suggested for such unsupervised learning of tuned units[24] (see also review in this issue by Abbott and Regehr, page 796). By contrast, the coefficients of the linear combination of the unit responses (labelled w_i in Box 1 and Fig. 3a), similar to the synaptic weights in neural networks, depend on the task and require at least some feedback about the output — for example, whether the 'male face' label was the correct answer or not. By definition, therefore, the modification of the weights during training is a supervised form of learning. The visual and motor tasks described in this paper are mostly supervised in the laboratory: for each example *x* (input), there is a label *y* (correct output); in experiment, monkeys receive feedback in every trial during training. Semi-supervised 'online' learning, however, in which feedback is provided for only some of the examples, is a better description of real-life visual and motor learning (see review in this issue by Tsodyks and Gilbert, page 775). Note that the full unsupervised learning problem (technically called density estimation) can be solved using supervised learning algorithms[56]. Furthermore, it turns out[58] that extending the regularization networks described in Box 1 to the unsupervised case is natural and does not change the basic architecture[62]. Biological learning is usually sequential and can therefore be characterized as online learning, in which the examples are provided one at a time. For online learning, biologically plausible versions of stochastic gradient descent can be used[25] (see review in this issue by Abbott and Regehr, page 796).

Because limbs are typically controlled by multiple sets of muscles (and an even larger number of muscle motor units), a major challenge in motor control has been to explain how the cortical cells modulate signals out of such large search space so that functional movements are generated. Previous work in vertebrates and invertebrates supports our hypothesis above, suggesting that specific motor behaviours are constructed through flexible combinations of a small number of modules, each generating a force field (in vertebrates a module is composed of a population of interneurons[40,41], but in invertebrates a single interneuron may function as a module[40]). According to this view, a module may reduce the number of degrees of freedom by controlling groups of muscles — and thus the assoc-iated field of forces — thereby functioning as a computational unit for use with different modulations in multiple motor behaviours[40,42,43]. Perhaps the most interesting aspect of the work was the discovery that the force fields induced by the focal activation of the cord follow a principle of linear combination[39,44] (see legend of Fig. 3 and Fig. 4), although this does not seem to hold for cats[45]). Specifically, Mussa-Ivaldi *et al.*[39] stimulated simultaneously two distinct sites in the frog's spinal cord and recorded the resultant forces at the ankle. They observed vector summation of the forces generated by each site separately: when the pattern of forces recorded at the ankle following co-stimulation were compared with those computed by summation of the two individual fields, they found that 'co-stimulation fields' and 'summation fields' were equivalent in more than 87% of cases. This is also true in the rat[46]. Moreover, the force-field summation underlies the control of limb trajectories in the frog[47].

Thus the hypothesis for explaining movement and posture is based on combinations of a few basic fields. The force fields (corresponding to the columns of the matrix *C* in the legend of Fig. 3) stored as synaptic weights in the spinal cord may be viewed as representing motor field primitives from which, through linear superimposition, a vast number of movements can be fashioned by impulses conveyed by supraspinal and reflex pathways. Computational analysis[48] verifies that this proposed mechanism is capable of learning and controlling a wide repertoire of motor behaviours.

Additional support to this view was provided by behavioural studies of reaching movements showing that when new forces are encountered, primates learn new dynamics to implement the desired trajectory[49]. Thoroughman and Shadmehr[50] were able to conclude from the pattern of generalization that the desired velocity of the reaching hand is mapped into a force required to move the hand at this velocity by combining tuned units with a gaussian shape. Their model can also be described in an equivalent, dual way as a combination of force fields (Fig. 3; Box 1).

In conclusion, there is independent evidence, in separate studies, for tuned neurons in motor cortex, and for a combination of a limited number of basic modules, each generating a force field and each modulated by supraspinal signals, in agreement with the caricature of Fig. 3b.

A canonical local circuit for tuning and generalization?

Thus, it seems that the combination of tuned receptive fields is the basic strategy used by both the visual and motor systems to learn and generalize. The similarity of the strategies in the visual and motor cortex, suggests that they might occur in other systems where learning is a component. The circuits that underlie the bell-shaped tuning curves are not known. Many cortical neurons seem to be tuned to a specific pattern of inputs, meaning that the maximum response of the cell occurs when the set of inputs takes specific activation values (which in general are not the set of maximum values of each input). It is a puzzle how this multidimensional tuning could be obtained parsimoniously by plausible neural circuits. One possibility is that tuning of a neuron to a specific set of activities of its many inputs (an infratemporal cortex neuron is likely to receive inputs from many cells, for instance from V4) is achieved by normalizing the inputs, which means dividing each one by the sum of the strengths of all of them. In fact, gaussian-like, multidimensional tuning — as found in many neurons in cortex —

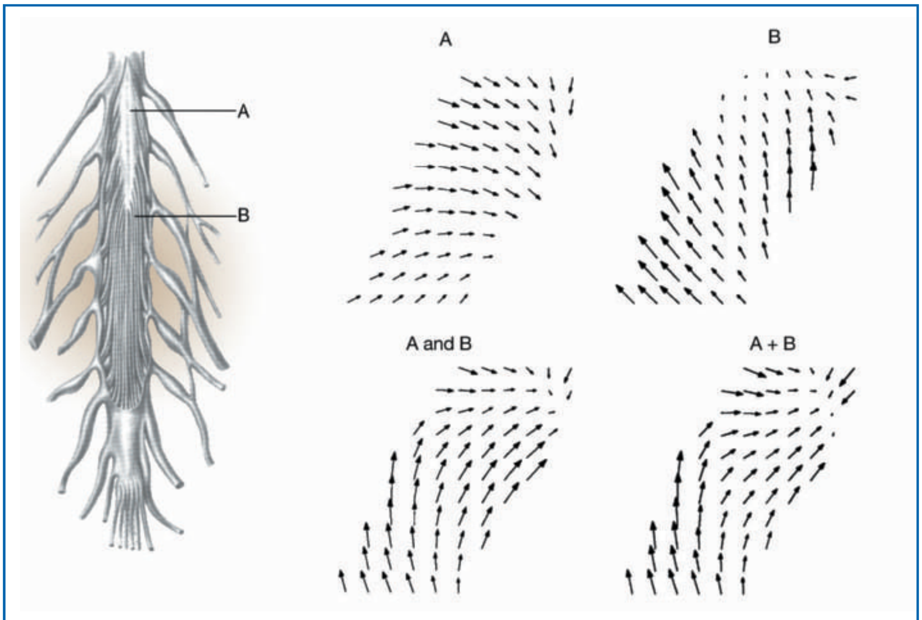


Figure 4 Spinal force fields combine linearly. Force fields A and B were obtained in response to stimulations delivered to two different spinal sites. The 'A and B' field was obtained by stimulating the same two sites simultaneously. It matches closely (correlation coefficient larger than 0.9) the force field A+B, which was calculated by pair-wise adding of the vector fields in A and in B. This highly linear behaviour applies to more than 87% of dual stimulation experiments. Adapted from Mussa-Ivaldi *et al.*[39].

can be generated by normalization of the input vector, followed by a simple threshold-like sigmoidal nonlinearity (Box 3).

Various neural circuits have been proposed to implement the key normalization stage, although the motivation behind the suggestions was to account for gain control and not tuning properties[51,52] (see review in this issue by Destexhe and Marder, page 789). Here, we propose that another role for normalizing local circuits in the brain is to provide (multidimensional) gaussian-shaped tuning, as a key step towards generalization. In fact, this might be the fundamental reason for the widespread presence of gain control circuits in cortex, where tuning to optimal stimuli is a common property. The normalization circuits might, for instance, use recurrent inhibition of the shunting type (Box 3), for which there is abundant evidence in cortex[53], although this is only one of several possibilities. Interestingly, the same basic circuit could implement the soft-max operation proposed for some of the processing stages in the visual system (Fig. 2). In any case, our new hypothesis is that gain control microcircuits underlie the tuning of cells to optimal stimuli in both the visual and motor systems.

Further questions in neuroscience and learning theory

Computational models versus experiments

Throughout this review, we used theoretical models as a tool to summarize experimental data provided by different approaches. The problems of visual recognition and motor control are computationally difficult and the experimental data from different sources are growing rapidly. We believe that quantitative models will increasingly replace the traditional qualitative mental models of the visual and motor physiologist and will become ever more important tools for interpreting data, and for planning and analysing experiments.

Time in vision and motor control

Our discussion of the visual system concentrated on the special case of recognition of a static image. In reality, we can recognize images that move and even sequences of movements. In the motor system, time has an even more obvious role: most of our motor commands deal with time-dependent motions and not simply with static postures. In vision, time can be introduced in a direct way assuming that visual neurons react to 'snapshots' of a motion and are selective for sequences of snapshots. In motor control, the equivalent assumption is that the motor primitives are time dependent. Box 1 suggests a strong analogy between vision and motor control in the time-dependent case: the basic strategy is to combine locally tuned units with time-dependent properties[54,55].

Hierarchical cortex architectures

It seems that modern learning theory does not offer any general argument in favour of hierarchical learning machines. This is a puzzle because the organization of cortex — as we argued for the visual and motor cortex — seems to be hierarchical.

Why hierarchies? There could be reasons of efficiency — computational speed and use of computational resources. For instance, the lowest levels of the hierarchy in visual cortex might represent a dictionary of features that can be shared across multiple classification tasks[56]. Hierarchical systems usually break down a task into a series of simple computations at each level. The same argument could apply to motor cortex. There might also be a more fundamental issue. Classical learning theory shows that the difficulty of a learning task depends on the complexity of the required learning architecture. This complexity determines in turn how many training examples are needed to achieve a given level of generalization. Thus, the complexity of the learning architecture sets the sample complexity for learning. If a task such as visual recognition can be decomposed into low-complexity learning tasks, for each layer of a hierarchical learning machine, then each layer might require only a small number of training examples. Of course, not all tasks have a hierarchical representation. Roughly speaking, the issue is about compositionality (S. Geman, personal communication): neuroscience suggests that what humans can learn — in vision and motor control — can be represented by hierarchies that are locally simple. Thus, our ability to learn from just a few examples, and its limitations, might be related to the hierarchical architecture of cortex.

Learning from very few examples

How then do the learning machines described in modern learning theory compare with brains? There are of course many aspects of biological learning that are not captured by the theory and several difficulties in making any comparison. One of the most obvious differences is the ability of people and animals to learn from very few examples. This is one of the challenges for the learning architectures we propose (although the networks described here, in particular the network of Fig. 2, can learn certain recognition tasks from less than ten labelled examples; M. Riesenhuber, personal communication). Of course, evolution has probably done a part of the learning and encoded it in the DNA. For instance, there is some evidence for basic face categorization ability to be present in human infants at birth, and for face-tuned neurons to be present in inferotemporal cortex of infant monkeys[57].

In any case, neuroscience suggests that an important area for future work on the theory and on the algorithms, is the problem of learning from partially labelled examples (and the related area of active learning): biological organisms usually have much visual and motor experience but mostly without direct or indirect feedback (providing the labels). Interesting theoretical work has begun on this; for example, showing that regularization networks (similar to the combination of tuned cells) could update their coefficients from a

Box 3

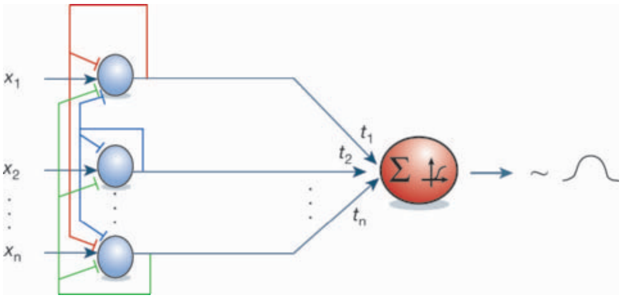
A neural circuit for gaussian-like tuning

For normalized t and x vectors (in the euclidean norm), the sigmoid of a scalar product can approximate a gaussian-like radial function[63]. Among the various neurally plausible circuits that have been proposed to approximate a normalization stage[51], we describe here a specific circuit, using lateral shunting inhibition[64–66], mainly to provide a possible example. There are certainly different possibilities for the nervous system to implement local normalization modules; for instance, using more complex synaptic properties (see review in this issue by Abbott and Regehr, page 796. The simplest equation — here in a time-independent form — describing a feedforward network of lateral shunting inhibition has the following form in a network of n cells:

$$y_i = \frac{h(x_i)}{c + \sum_k k(x_k)} \qquad i = 1, \ldots, n.$$

where h and k represent the transduction between nonlinear presynaptic and postsynaptic voltage at the output of cell i and at the output of the interneurons mediating lateral inhibition, respectively. If $h(x) = x$ and $k(x) \cong \sqrt{x^2}$, the circuit performs a normalization operation; if $h(x) \cong x^{q+1}$ and $k(x) \cong x^q$ with q sufficiently large ($q \geqslant 2$), then the circuit performs a max operation, for example, $y_i \cong x_i$ if $x_i = \max_j x_j$, otherwise $y_i \cong 0$ (see ref. 67). The figure shows the circuit with inhibition of the shunting type (the arrows indicate depolarizing synapses, whereas the symbol ⊣ indicates shunting inhibition onto interneurons (blue). Depending on the parameters, the activity of the tuned output cell (red) — after summation of the inputs with x_1, x_2, ..., x_n weighted with synaptic weights $t_1, t_2, \ldots, t_n$ and then transformation through a sigmoidal, threshold-like nonlinearity, such as provided by the spike mechanism — can approximate a gaussian-like, bell-shaped function of the inputs, that is $e^{-(x_1-t_1)^2-(x_2-t_2)^2 \ldots -(x_n-t_n)^2}$, since the input vector is normalized by the recurrent inhibitory circuit.

Note that the neuron responds maximally to the 'optimal' pattern of inputs with values $t_1, t_2, \ldots, t_n$. Note also that the same basic circuit of lateral inhibition with somewhat different synaptic parameters could underlie gaussian-like tuning (by means of normalization) and the softmax operation[54] — which are the two key operations required at various stages in the model of object recognition shown in Fig. 2.



partially labelled set of examples[58]. Other approaches, such as bayesian and graphical models, might be able to deal more generally with the problem of unsupervised learning (for example, ref. 25).

The mind as a theory of the world

In modern mathematical theory, the property of generalization is the key property of learning. Learning, as opposed to memory, synthesizes functions that are predictive of the world. Thus, learning synthesizes modules — such as vision and motor control — that are effectively theories of the physical world, in the sense of being predictive of specific aspects of it. Learning is done within these architectures by the plasticity of synapses, and learning is what makes the brain a theory of the world.

The quest for generalization mechanisms

There is considerable evidence in the visual and motor system for the learning architectures we propose — a combination of tuned units. But whether our hypothesis is a satisfactory, first-order description of the first few hundred milliseconds of visual perception and motor control or whether more complex, recurrent network models will be needed remains unclear. It will be interesting to look at other systems, such as the auditory, somatosensory and olfactory systems, from a similar point of view. There is little evidence at this point for or against our proposal of a canonical microcircuit underlying tuning in many neurons throughout the brain[52,59]; there is even less evidence for the specific circuit we suggest. In fact, other plausible neural and synaptic circuits could work as well. Finally, it is unclear whether similar, simple learning architectures could have any role in typical human brain functions such as learning language. □

doi.10.1038/nature03014

1. Vapnik, V. N. *Statistical Learning Theory* (Wiley, New York, 1998).
2. Bülthoff, H. & Edelman, S. Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proc. Natl Acad. Sci. USA* **89,** 60–64 (1992).
3. Riesenhuber, M. & Poggio, T. Hierarchical models of object recognition in cortex. *Nature Neurosci.* **2,** 1019–1025 (1999).
4. Riesenhuber, M. & Poggio, T. in *The Visual Neurosciences* Vol. 2 (eds Chalupa, L. M. & Werner, J. S.) 1640–1653 (MIT Press, Cambridge, MA, 2003).
5. Palmeri, T. & Gauthier, I. Visual object understanding. *Nature Rev. Neurosci.* **5,** 291–303 (2004).
6. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **160,** 106–154 (1962).
7. Hubel, D. & Wiesel, T. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *J. Neurophysiol.* **28,** 229–289 (1965).
8. Mountcastle, V. B. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* **20,** 408–434 (1957).
9. Gross, C. G. in *Handbook of Sensory Physiology* Vol. VII/3B (eds Autrum, H., Jung, R., Lowenstein, W., Mckay, D. & Teuber, H.-L.) (Springer, Berlin, 1973).
10. Bruce, C., Desimone, R. & Gross, C. Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque. *J. Neurophysiol.* **46,** 369–384 (1981).
11. Perrett, D. *et al.* Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Exp. Brain Res.* **86,** 159–173 (1991).
12. Perrett, D. & Oram, M. Neurophysiology of shape processing. *Img. Vis. Comput.* **11,** 317–333 (1993).
13. Logothetis, N., Pauls, J. & Poggio, T. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* **5,** 552–563 (1995).
14. Logothetis, N. & Sheinberg, D. Visual object recognition. *Annu. Rev. Neurosci.* **19,** 577–621 (1996).
15. Kobatake, E. & Tanaka, K. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *J. Neurophysiol.* **71,** 856–857 (1994).
16. Kobatake, E., Wang, G. & Tanaka, K. Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *J. Neurophysiol.* **80,** 324–330 (1998).
17. DiCarlo, J. & Maunsell, J. Form representation in monkey inferotemporal cortex is virtually unaltered by free viewing. *Nature Neurosci.* **3,** 814–821 (2000).
18. Booth, M. & Rolls, E. View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb. Cortex* **8,** 510–523 (1998).
19. Tanaka, K. Neuronal mechanisms of object recognition. *Science* **262,** 685–688 (1993).
20. Sato, T. Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake monkeys. *Exp. Brain Res.* **77,** 23–30 (1989).
21. Hietanen, J., Perrett, D., Benson, P. & Dittrich, W. The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp. Brain Res.* **89,** 157–171 (1992).
22. Missal, M., Vogels, R. & Orban, G. Responses of macaque inferior temporal neurons to overlapping shapes. *Cereb. Cortex* **7,** 758–767 (1997).
23. Poggio, T. A. Theory of how the brain might work. *Cold Spring Harb. Symp. Quant. Biol.* **4,** 899–910 (1990).
24. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381,** 607–609 (1996).
25. Pouget, A., Dayan, P. & Zemel, R. S. Computation and inference with population codes. *Annu. Rev. Neurosci.* **26,** 381–410 (2003).
26. Pouget, A. & Sejnowski, T. J. Spatial transformations in the parietal cortex using basis functions. *J. Cogn. Neurosci.* **9,** 222–237 (1997).
27. 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).
28. Amirikian, B. & Georgopoulos, A. P. Modular organization of directionally tuned cells in the motor cortex: is there a short-range order? *Proc. Natl Acad. Sci. USA* **100,** 12474–12479 (2003).
29. Merzenich, M. M. & Brugge, J. F. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. *Brain Res.* **50,** 275–296 (1973).
30. Wise, S. P., Moody, S. L., Blomstrom, K. J. & Mitz, A. R. Changes in motor cortical activity during visuomotor adaptation. *Exp. Brain Res.* **121,** 285–299 (1998).
31. Paz, R., Boraud, T., Natan, C., Bergman, H. & Vaadia, E. Preparatory activity in motor cortex reflects learning of local visuomotor skills. *Nature Neurosci.* **6,** 882–890 (2003).
32. Gribble, P. L. & Scott, S. H. Overlap of internal models in motor cortex for mechanical loads during reaching. *Nature* **417,** 938–941 (2002).
33. Gandolfo, F., Li, C. R., Benda, B., Padoa-Schioppa, C. & Bizzi, E. Cortical correlates of motor learning in monkeys adapting to a new dynamic environment. *Proc. Natl Acad. Sci. USA* **97,** 2259–2263 (2000).

34. Li, C. R., Padoa-Schioppa, C. & Bizzi, E. Neuronal correlates of motor performance and motor learning in the primary motor cortex of monkeys adapting to an external force field. *Neuron* **30,** 593–607 (2001).
35. Grillner, S. & Wallen, P. Central pattern generators for locomotion, with special reference to vertebrates. *Annu. Rev. Neurosci.* **8,** 233–261 (1985).
36. Stein, P. S., Victor, J. C., Field, E. C. & Currie, S. N. Bilateral control of hindlimb scratching in the spinal turtle: contralateral spinal circuitry contributes to the normal ipsilateral motor pattern of fictive rostral scratching. *J. Neurosci.* **15,** 4343–4355 (1995).
37. Loeb, G. E. Motoneurone task groups: coping with kinematic heterogeneity. *J. Exp. Biol.* **115,** 137–146 (1985).
38. Bizzi, E., Giszter, S. & Mussa-Ivaldi, F. A. Computations underlying the execution of movement: a novel biological perspective. *Science* **253,** 287–291 (1991).
39. Mussa-Ivaldi, F. A., Giszter, S. F. and Bizzi, E. Linear combinations of primitives in vertebrate motor control. *Proc. Natl Acad. Sci. USA* **91,** 7534–7538 (1994).
40. Jing, J., Cropper, E. C., Hurwitz, I. & Weiss, K. R. The construction of movement with behavior-specific and behavior-independent modules. *J. Neurosci.* **24,** 6315–6325 (2004).
41. Saltiel, P., Tresch, M. C. & Bizzi, E. Spinal cord modular organization and rhythm generation: an NMDA iontophoretic study in the frog. *J. Neurophysiol.* **80,** 2323–2339 (1998).
42. Grillner, S. in *Handbook of Physiology — The Nervous System* edn 4 (eds Brookhart, J. M. & Mountcastle, V. B.) 1179–1236 (American Physiological Society, Bethesda, MD, 1981).
43. d'Avella, A., Saltiel, P. & Bizzi, E. Combinations of muscle synergies in the construction of a natural motor behaviour. *Nature Neurosci.* **6,** 300–308 (2003).
44. Lemay, M. A., Galagan, J. E., Hogan, N. & Bizzi, E. Modulation and vectorial summation of the spinalized frog's hindlimb end-point force produced by intraspinal electrical stimulation of the cord. *IEEE Trans. Neural. Syst. Rehabil. Eng.* **9,** 12–23 (2001).
45. Aoyagi, Y., Stein, R. B., Mushahwar, V. K. & Prochazka, A. The role of neuromuscular properties in determining the end-point of a movement. *IEEE Trans. Neural. Syst. Rehabil. Eng.* **12,** 12–23 (2004).
46. Tresch, M. C. & Bizzi, E. Responses from the spinal microstimulation in the chronically spinalized rats and their relationship to spinal systems activated by low threshold cutaneous stimulation. *Exp. Brain Res.* **129,** 401–416 (1999).
47. Kargo, W. J. & Giszter, S. F. Rapid correction of aimed movements by summation of force field primitives. *J. Neurosci.* **20,** 409–426 (2000).
48. Mussa-Ivaldi, F. A. in *Proc. 1997 IEEE Int. Symp. Computational Intelligence in Robotics and Automation* 84–90 (IEEE Computer Society, Los Alamitos, California, 1997).
49. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14,** 3208–3224 (1994).
51. Chance, F., Nelson, S. & Abbott, L. Complex cells as cortically amplified simple cells. *Nature Neurosci.* **2,** 277–282 (1999).
50. Thoroughman, K. & Shadmer, R. Learning of action through adaptive combination of motor primitives. *Nature* **407,** 740–746 (2000).
52. Douglas, R. & Martin, K. A functional microcircuit for cat visual cortex. *J. Physiol. (Lond.)* **440,** 735–769 (1991).
53. Borg-Graham, L. J., Monier, C. & Frégnac Y. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393,** 369–373 (1998).
54. Mussa-Ivaldi, F. A. & Bizzi, E. Motor learning through the combination of primitives. *Phil. Trans. R. Soc. Lond. B* **355,** 1755–1769 (2000).
55. Giese, M. & Poggio, T. Neural mechanisms for the recognition of biological movements. *Nature Rev. Neurosci.* **4,** 179–192 (2003).
56. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* (Springer, Basel, 2001).
57. Rodman, H. R., Scalaidhe, S. P. & Gross, C. G. Response properties of neurons in temporal cortical visual areas of infant monkeys. *J. Neurophysiol.* **70,** 1115–1136 (1993).
58. Belkin, M., Niyogi, P. & Sindhwani, V. Technical Report TR–2004–05 (University of Chicago, Chicago, 2004).
59. Douglas, R. & Martin, K. Neural circuits of the neocortex. *Annu. Rev. Neurosci.* **27,** 419–451 (2004).
60. Salinas, E. & Abbott, L. F. Transfer of coded information from sensory to motor networks. *J. Neurosci.* **15,** 6461–6474 (1995).
61. Hogan, N. An organizing principle for a class of voluntary movements. *J. Neusosci.* **4,** 2745–2754 (1984).
62. Poggio, T. & Smale, S. The mathematics of learning: dealing with data. *Notices Am. Math. Soc.* **50,** 537–544 (2003).
63. Maruyama, M., Girosi, F. & T. Poggio, T. A. *Connection Between GRBF and MLP*. AI Memo 1291 (Massachusetts Institute of Technology, Cambridge, Massachusetts 1992).
64. Carandini, M. & Heeger, D. J. Summation and division by neurons in visual cortex. *Science* **264,** 1333–1336 (1994).
65. Carandini, M., Heeger, D. J. & Movshon, J. A. Linearity and normalization in simple cells of the macaque primary visual cortex. *J. Neurosci.* **17,** 8621–8644 (1997).
66. Reichardt, W., Poggio, T. & Hausen, K. Figure-ground discrimination by relative movement in the visual system of the fly II: towards the neural circuitry. *Biol. Cybern.* **46,** 1–30 (1983).
67. Yu, A. J., Giese, M. A. & Poggio, T. Biophysiologically plausible implementations of the maximum operation. *Neural Comput.* **14,** 2857–2881 (2002).

Acknowledgements We thank S. Mussa-Ivaldi, R. Shadmeher, G. Kreiman, and M. Riesenhuber for insightful and helpful comments. This research was sponsored by grants from NIH, Office of Naval Research, DARPA and National Science Foundation. Additional support was provided by Eastman Kodak Company, Daimler Chrysler, Honda Research Institute, NEC Fund, Siemens Corporate Research, Toyota, Sony and the McDermott chair (T.P.).

Competing interests statement The authors declare that they have no competing financial interests.