

A Neural Model of Perceptual-Motor Alignment

Emmanuel Guigon¹ and Pierre Baraduc²

Abstract

■ Sensorimotor systems face complex and frequent discrepancies among spatial modalities, for example, growth, optical distortion, and telemanipulation. Adaptive mechanisms must act continuously to restore perceptual-motor alignments necessary for perception of a coherent world. Experimental manipulations that exposed participants to localized discrepancies showed that adaptation is revealed by the acquisition of a constrained relation between entire modalities rather than associations between individual exemplars within these modalities. The computational problem faced by the human nervous system can thus be conceived as having to induce constrained

relations between continuous stimulus and response dimensions from ambiguous or incomplete training sets, that is, performing interpolation and extrapolation. How biological neuronal networks solve this problem is unknown. Here we show that neural processing based on linear collective computation and least-square (LS) error learning in populations of frequency-coded neurons (i.e., whose discharge varies in a monotonic fashion with a parameter) has built-in interpolation and extrapolation capacities. This model can account for the properties of perceptual-motor adaptations in sensorimotor systems. ■

INTRODUCTION

Perceptual-motor distortions have long been used to explore adaptive capacities of sensorimotor systems (Redding & Wallace, 1997; Welch, 1978; Held, 1965). In a typical experiment, participants look through an optical device that displaces the visual field laterally (prism), and try to reach visual targets. In this condition, adaptation is revealed by specific behavioral modifications during exposure (visuomotor reduction of effect) and following removal of the prism (visuomotor negative aftereffect) (Welch, 1986). More generally, human subjects can adapt more or less completely to many types of distorted environments and distorting devices acting on different modalities (Welch, 1986). Much has been learned about the conditions that produce adaptation, the nature of adaptation, and how adaptation modifies behavior (Redding & Wallace, 1997; Welch, 1986), and a detailed conceptual and functional model has recently been proposed to account for prism adaptation (Redding & Wallace, 1997). Nevertheless, no mechanism has yet been described that could explain how changes in neural operations resulting from exposure to sensorimotor discrepancies actually define an adapted behavior.

The nature of the sought-after mechanism is constrained by two cardinal features of perceptual-motor adaptations (Vetter, Goodbody, & Wolpert, 1999; Shinn-

Cunningham, Durlach, & Held, 1998; Schor, Gleason, Maxwell, & Lunn, 1993; Bedford, 1989, 1993a, 1993b; Hay, 1974). First, adapted behavior is not a collection of input/output associations, but a true relation between entire dimensions of stimuli. Second, a linear constraint restricts the range of possible adaptations. These properties have been revealed by experimental manipulations that exposed participants to localized discrepancies. For example, Bedford (1989) studied the acquisition of new visual-proprioceptive mappings specified by a small number of isolated pairs of visually and proprioceptively felt positions. She found that training at a single position led to a global realignment, and training at two or three positions resulted in a linear mapping that resembled the least-square (LS) approximation of the training set (Bedford, 1989). Closely related observations were made in experiments using other modalities (Shinn-Cunningham et al., 1998; Schor et al., 1993).

The problem faced by the nervous system in these circumstances is computationally “ill posed,” and requires rules to interpolate between the training pairs and extrapolate outside the training range. It can be considered in the framework of “function approximation” (Girosi, Jones, & Poggio, 1995; Poggio & Girosi, 1990) and cast as a problem of finding a set of “basis functions” whose interpolation and extrapolation capacities meet the preceding requirements. The goal is not to find the most efficient kind of approximating functions (Girosi et al., 1995; Hornik, Stinchcombe, & White, 1989), but functions that match neural constraints. Two types of basis functions are classically used in neural modeling (Pouget & Sejnowski, 1994): radially

¹INSERM U483, Université Pierre et Marie Curie, ²University College London

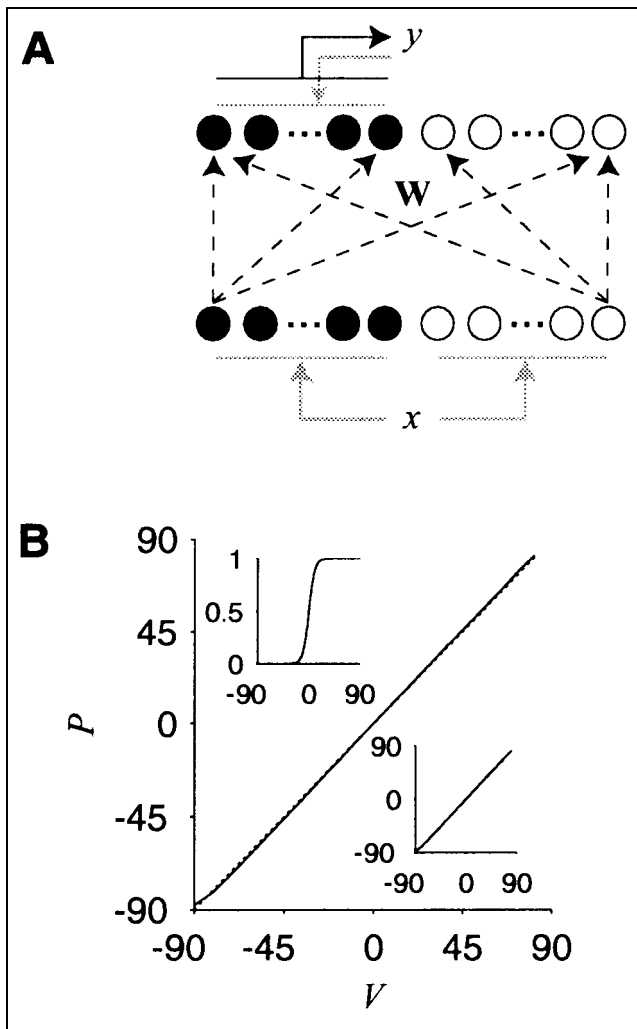


Figure 1. (A) Mapping between dimensions x and y through a fully connected feedforward network: positive populations (\bullet), negative populations (\circ), encoding (gray line), decoding (dark line), modifiable connections (dashed line). (B) Noise-free performance of the decoding method on the positive output population following training (100,000 iterations, $\eta = 0.0005$) on the identity mapping. Dashed line corresponds to a perfect mapping. Lower inset shows the noise-free performance of the decoding method on the positive input population. Upper inset shows the response function of a single neuron.

symmetric functions (e.g., Gaussians) to represent spatial variables defined by receptive fields and vectorial variables, and “sigmoid functions” to encode values of intensity parameters (e.g., interaural intensity difference) or postural parameters that are derived from proprioceptive and efferent copy sources (Baraduc, Guigon, & Burnod, 2001; Salinas & Abbott, 1995; Pouget & Sejnowski, 1994; Dean, 1990; Olson & Hanson, 1990; Zipser & Andersen, 1988). The above mentioned adaptation experiments use egocentric localization parameters (e.g., gaze direction, hand position, vergence, interaural intensity difference) that follow the latter coding scheme (see Discussion). In this article, we show that the properties of perceptual-motor adaptations can

be explained by linear neural computation based on sigmoid functions (Figure 1) (and more generally on monotonic functions), but not radial functions. In the case of 1-D parameters, this result is proven mathematically for linear functions and with numerical simulations for nonlinear functions. A tentative extension to 2-D adaptation paradigms (e.g., Ghahramani, Wolpert, & Jordan, 1996) is then proposed.

RESULTS

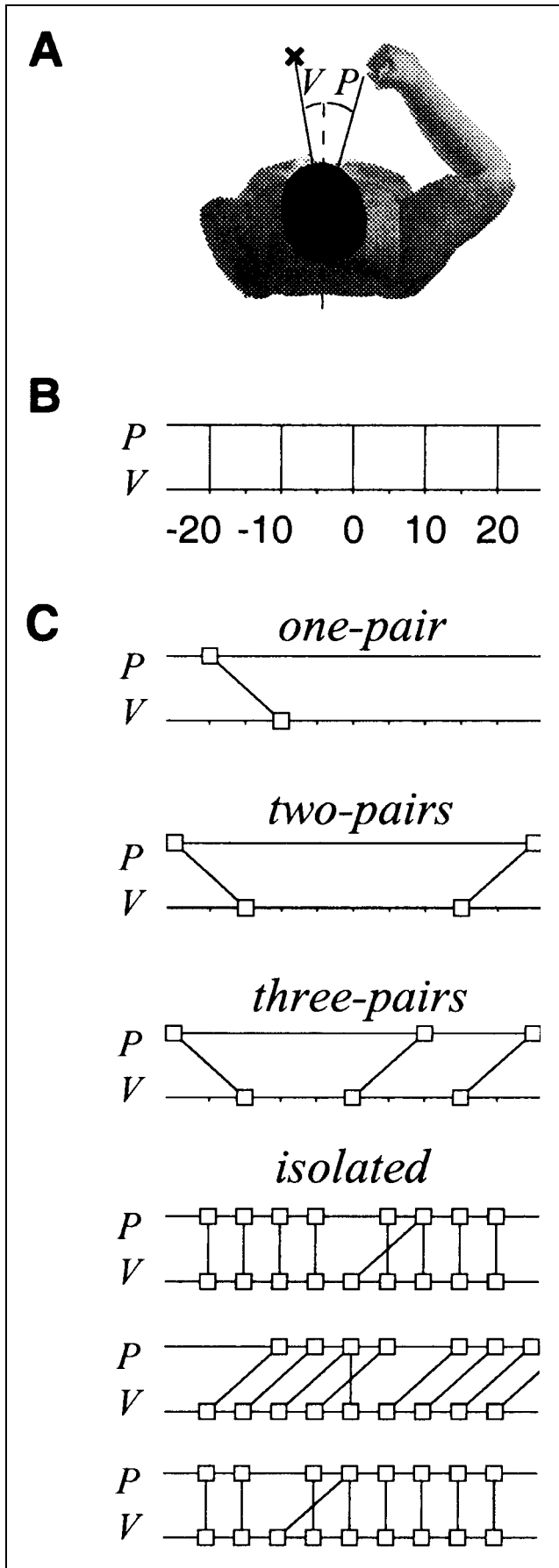
1-D Visual-Proprioceptive Adaptation

A single perturbation (one-pair experiment; Figure 2C) led to a global generalization to untrained values of the input dimension (Figure 3A). The result is shown for a leftward position, but similar results were obtained for all positions. The two-pair experiment produced a linear mapping between the training pairs ($P = 0.42V - 0.05$, $R^2 = .997$) and a flattening beyond the training range (Figure 3B). A similar result was obtained for opposite direction offsets ($-15 \rightarrow -15 + 10$, $15 \rightarrow 15 - 10$) (not shown). When a third nonaligned pair was introduced (three pairs), the network still generated a linear mapping (Figure 3C). Its slope was similar to that of the two-pair experiment, but the intercept changed to account for the new training pair ($P = 0.42V + 3.55$, $R^2 = .988$). The linear trend of the training set was $P = 0.67V$ and $P = 0.67V + 3.33$ in the two- and three-pair experiments, respectively.

We explored extrapolation properties of the model by manipulations of the two-pair experiment (Bedford, 1993b). First, we observed that extrapolation failed to occur for a smaller offset (Figure 4A). This observation was true for any offset, showing that the absence of extrapolation was not due to a performance limitation related to the size of the offset. Second, linear interpolation was not restricted to the central straight-ahead region, but occurred within the training range in non-central positions as well (Figure 4B). Third, extrapolation in a range of responses that is smaller than that encountered during training is similar to that found for responses larger than already encountered (Figure 4C). Fourth, no generalization decrement was found when one of the pairs had no offset (Figure 4D). In this latter case, decrement around the distorted position might have been a more conservative strategy.

An open question was whether stronger training constraints could have forced the network to single out a training position. This issue was addressed with the “isolated” experiment (Figure 2C). In fact, the network was unable to learn an isolated training pair correctly when explicitly instructed to do so (Figure 5). Instead, a rigid shift was observed just as in the one-pair experiment.

We next asked whether we could distinguish trained from untrained positions in the adapted network. To

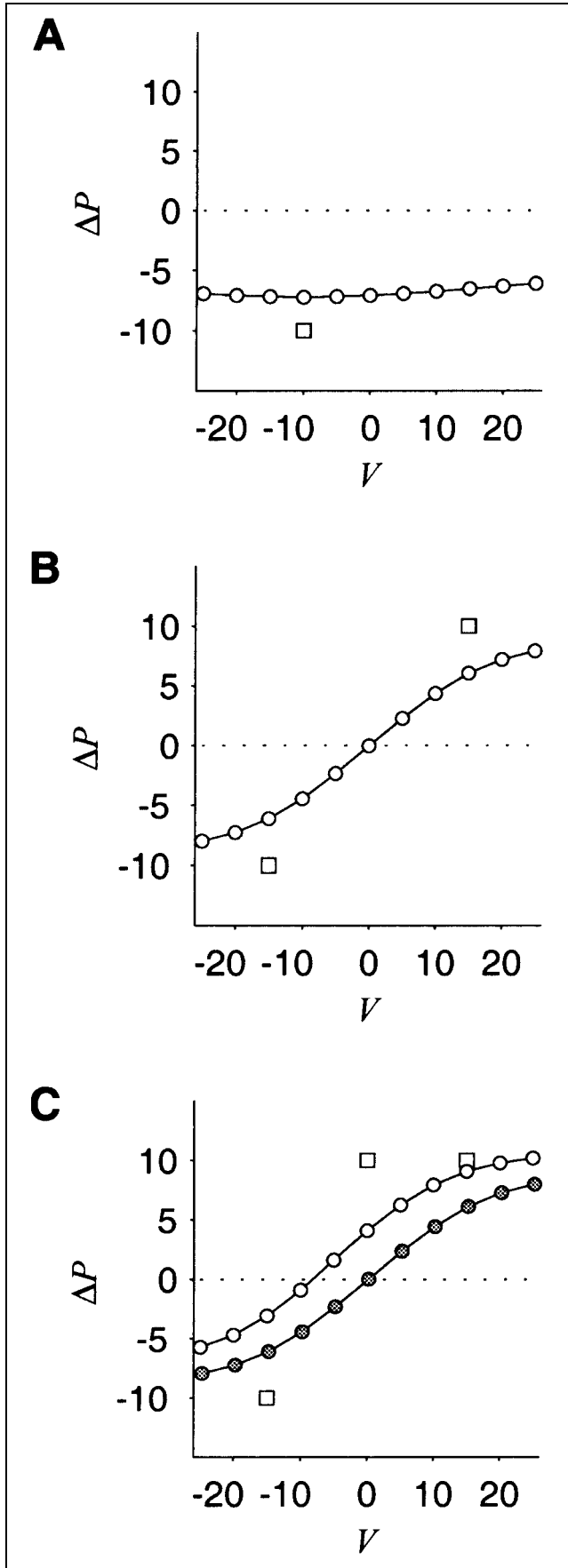


this end, we calculated the variability of the network response, that is, the variance of network output when the input activity profile was corrupted by additive Gaussian noise. We used a two-pair adaptation task as in Bedford (1993a). This variance was uniform across the input dimension (Figure 6). Thus, despite the fact that the network was trained to memorize specific exemplars, these became indistinguishable following training: The emergent behavior of the model was not the consequence of the training set alone, but also reflected intrinsic characteristics of the computation.

Altogether these results are in close agreement with adaptive properties of the human visual-proprioceptive system (Bedford, 1989, 1993a, 1993b). For each experiment, however, the agreement is true for certain numbers of training blocks, but need not hold in general when learning reaches asymptote. In this latter case, nonlinear mappings develop that provide better fits to the training sets (Figure 7). Steady-state adaptation has not been studied experimentally, and it is an open question as to whether participants could acquire nonlinear mappings, and whether these mappings would resemble those predicted by the model.

The preceding results were obtained with a value of s that corresponded to a compromise between two extreme cases. On the one hand, when s was small, the response functions could be approximated by Heaviside functions, and they could be linearly combined to represent any simple function (i.e., constant on intervals). Thus, in the limit of a large number of neurons, the network could theoretically realize any integrable function. In this case, sigmoidal generalization was observed in a two-pair experiment (Figure 8A). On the other hand, when s was large, the response functions were approximately linear, and a linear combination of these functions was close to linear (Figure 8B). In fact, in the strictly linear case, it could be shown that LS error learning in the network is mathematically equivalent to the LS approximation of the training set (Appendix B). Whenever the neurons encountered restricted nonlinearities in the task range, the mapping remained close to the LS approximation of the training set for long training periods because the development of a better-fitting nonlinear mapping was determined by the degree of nonlinearities in the response functions.

Figure 2. (A) Visual-proprioceptive transformation. Target (cross) and hand locations are measured by the egocentric angles V and P , respectively. (B) Alignment between V and P . Scale is in degrees. (C) Four types of transformation of the alignment between V and P . A new mapping is specified by a restricted number of input/output pairs (\square). For instance, in the one-pair experiment, pointing toward a target at -10° required a movement toward -20° . No information was given for other targets. In the isolated experiments, the central training pair was presented on half of the trials.



The specificity of these results is illustrated by a comparison with a Gaussian model. Simulations were run with Gaussian encoding,

$$x_i = \exp\left(-\frac{(x - \lambda_i)^2}{2\sigma^2}\right)$$

where σ is the width of the Gaussian. Center-of-mass decoding (Snippe, 1996) was used for narrow tuning curves and LS decoding for broad tuning curves. Network architecture and functioning were as for the sigmoid model, but with a single input, and output populations each containing $N = 100$ neurons. Parameter σ was either 8.4° (narrow Gaussian model) or 47.6° (broad Gaussian model). These values corresponded to a half-width at half-height of 10° and 56° , respectively, falling below the minimum and in the median range of tuning widths of motor cortical cells, respectively (Amirikian & Georgopoulos, 2000).

Generalization was characterized by localized changes around the training positions (Figure 8C and D). The inferred mapping was monotonic in Figure 8D, but was clearly different from the mappings discovered by the monotonic network. In particular, it tended to return toward the initial mapping outside the training range. These results show that a Gaussian model is unable to explain adaptive properties of the human visual-proprioceptive system, even with broadly tuned computational elements.

2-D Visual-Proprioceptive Adaptation

The model can be immediately extended to the multi-dimensional case by using monotonic surfaces as response functions. An open issue is the nature of the response surfaces that could make the model compatible with adaptation experiments in a 2-D workspace (Ghahramani et al., 1996). We will not address this general question directly, but will simply show that linear saturated functions are appropriate to this task.

For the sake of simplicity, we considered first the case of linear response functions, where no neuron encounters a nonlinearity in the task range. We further assumed that the input and output dimensions were described in Cartesian coordinates, so we did not address the problem of coordinate transformations (Vetter et al., 1999; Ghahramani et al., 1996). Encoding in the input layer was defined by the affine mapping $\mathcal{C}^{\text{in}}: \mathbb{R}^2 \mapsto \mathbb{R}^N$, $x \rightarrow \mathcal{C}^{\text{in}}(x) = \mathcal{M}^{\text{in}} x + E^{\text{in}}$, where \mathcal{M}^{in} is an $N \times 2$ matrix and $E^{\text{in}} \in \mathbb{R}^N$ (N is the number of neurons). In the same way, encoding in the output layer was defined by the affine

Figure 3. Perceptual-motor alignment for the sigmoid model. Training pairs are depicted by \square . (A) One-pair experiment, 50 blocks. (B) Two-pairs experiment, 300 blocks. (C) Three-pairs experiment, 300 blocks. Gray circles indicate the result obtained in (B).

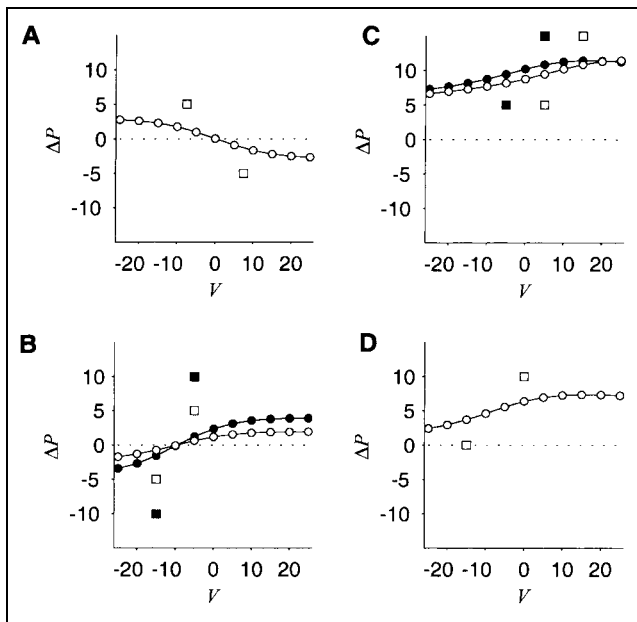


Figure 4. Extrapolation behavior of the model. Variations of the two-pairs experiment, 300 blocks. (B) and (C) show the results of two experiments (open circle for open square targets, closed circle for closed square targets).

mapping $\mathcal{C}^{\text{out}}: \mathbb{R}^2 \mapsto \mathbb{R}^N, x \mapsto \mathcal{C}^{\text{out}}(x) = \mathcal{M}^{\text{out}}x + E^{\text{out}}$. Decoding the output layer was defined by $\mathcal{D}^{\text{out}}: \mathbb{R}^N \mapsto \mathbb{R}, X \mapsto \mathcal{D}^{\text{out}}(X) = (\mathcal{M}^{\text{out}})^*(X - E^{\text{out}})$, where $(\mathcal{M}^{\text{out}})^*$ is the left pseudo-inverse of \mathcal{M}^{out} . The task workspace was the unit square. The tuning parameters ($\mathcal{M}^{\text{in}}, E^{\text{in}}, \mathcal{M}^{\text{out}}, E^{\text{out}}$; same for input and output) were chosen to cover the workspace approximately uniformly.

Adaptation at the central position led to a uniform generalization to the whole workspace (Figure 9A). Adaptation to two opposite displacements at two different positions led to a uniform generalization along the direction of displacement, and a linear generalization in the perpendicular direction (Figure 9B). For the latter configuration, we also calculated the pattern of adaptation predicted by the LS approximation of the training pairs (Figure 9C). This pattern did not resemble the outcome of the model or the experimental observations (Ghahramani et al., 1996). The proof in Appendix B does not extend to the multidimensional case.

We replicated these experiments with nonlinear response functions. These functions were similar to those used in the linear case, but with lower and upper saturations (at 0 and 1). Decoding was performed by searching for the output coordinates that best predicted (in the LS sense) the observed activity profile. The results (Figure 9D and E) are similar to those obtained in the linear case (Figure 9A and B).

Vetter et al. (1999) studied adaptation to a distortion at a single noncentral position in a frontal plane. We simulated this case by testing adaptation in a restricted central region of the workspace with a distorted point

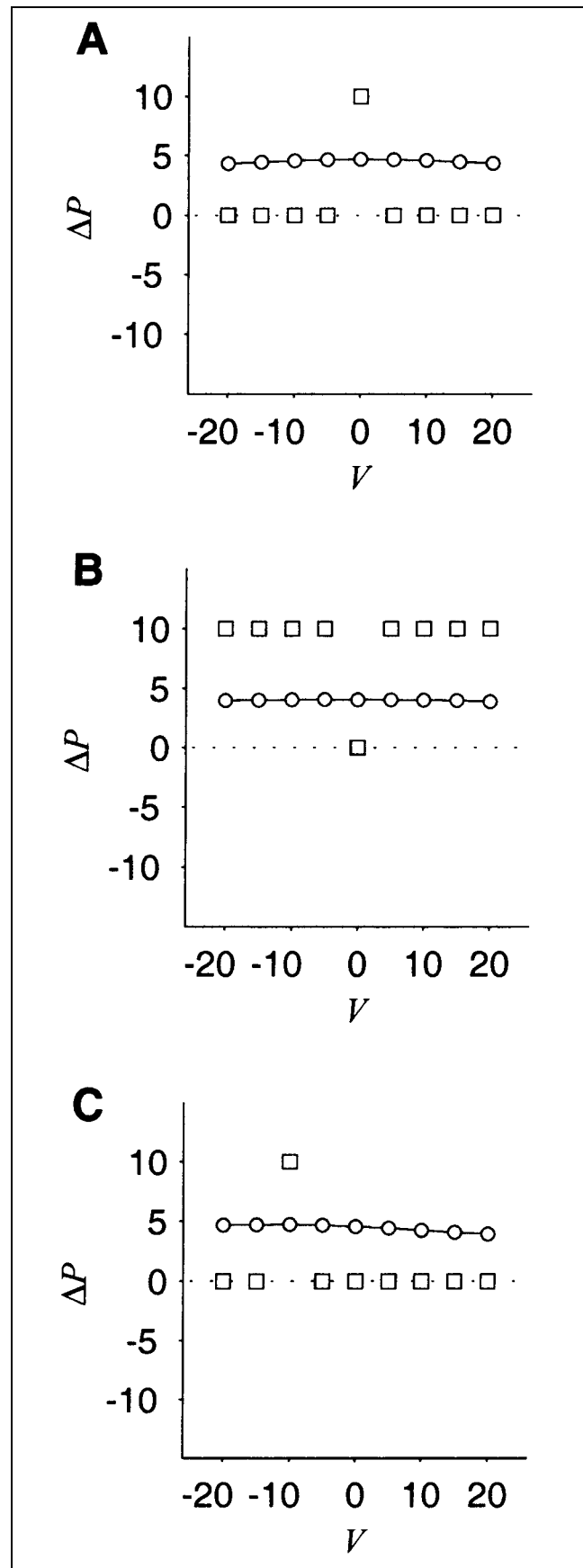


Figure 5. Isolated experiments, five blocks. See Figure 2C and its legend.

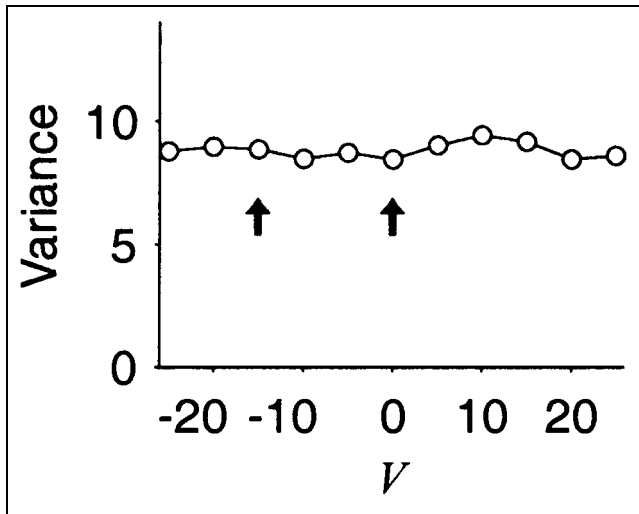


Figure 6. Output variability measured following injection of additive Gaussian noise (mean = 0, $SD = 0.1$, 2000 repetitions) in the input layer in a two-pairs experiment (arrows, $V = -15 \rightarrow P = V$ and $V = 0 \rightarrow P = V + 10$).

at the border of this region. The resulting pattern resembled the abovementioned two-point case, with a uniform generalization along the direction of the distortion and a linear generalization along the other axis (Figure 9A). The slope of this linear trend depended on the position of the distorted point along the vertical axis (on the figure) and the amplitude of the distortion. We calculated the average magnitude of adaptation as a function of the distance to the exposure point. The magnitude decreased linearly with distance (Figure 9F). Vetter et al. (1999) made a similar calculation over a 3-D region and reported a statistically nonsignificant

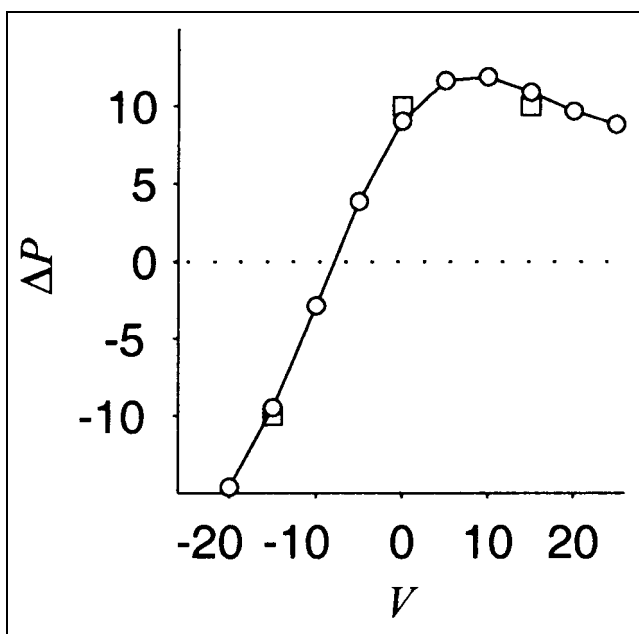


Figure 7. Three-pairs experiment, 5000 blocks.

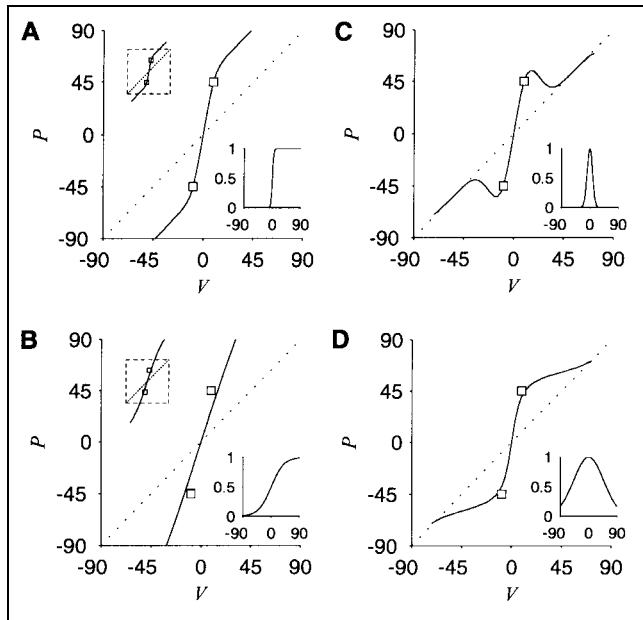


Figure 8. Generalization behavior of monotonic and Gaussian populations. Training pairs are depicted by \square : $(-9; -45)$ and $(9; 45)$. Each pair was presented 2000 times. Learning rate was 0.0005. Inset in the lower part of the plots depicts the actual response function. Inset in the upper part depicts the behavior of the network outside the unit square (dashed line). (A) Monotonic populations, $s = 2$. (B) Monotonic populations, $s = 20$. (C) Gaussian populations, $\sigma = 8.4$. (D) Gaussian populations, $\sigma = 47.6$.

decreasing trend. The distance effect was not apparent in the nonlinear case (Figure 9F).

DISCUSSION

The idea that an infinite number of input/output mappings are compatible with only a small number of input and output pairs was originally used to understand how human subjects (e.g., experts) infer laws from a restricted set of examples (“function learning”; Brehmer, 1974; Carroll, 1963). When applied to the sensorimotor system, this principle helps reveal internal rules that guide the formation of mappings between sensory and motor dimensions (Bedford, 1989). Here, we addressed these rules by exploring interpolation and extrapolation capacities of single-layer linear neural networks. We showed that linear computations between populations of sigmoid neurons explain two salient features of perceptual-motor adaptations: (1) Adaptation is characterized by a relation between entire dimensions of stimuli. (2) This relation is shaped by a linearity constraint.

The sigmoid coding scheme has been considered, together with the Gaussian scheme, in the field of function approximation (Girosi et al., 1995; Hornik et al., 1989). In this framework, families of sigmoid and Gaussian functions have been attributed universal approximation capacities (Girosi et al., 1995; Hornik et al.,

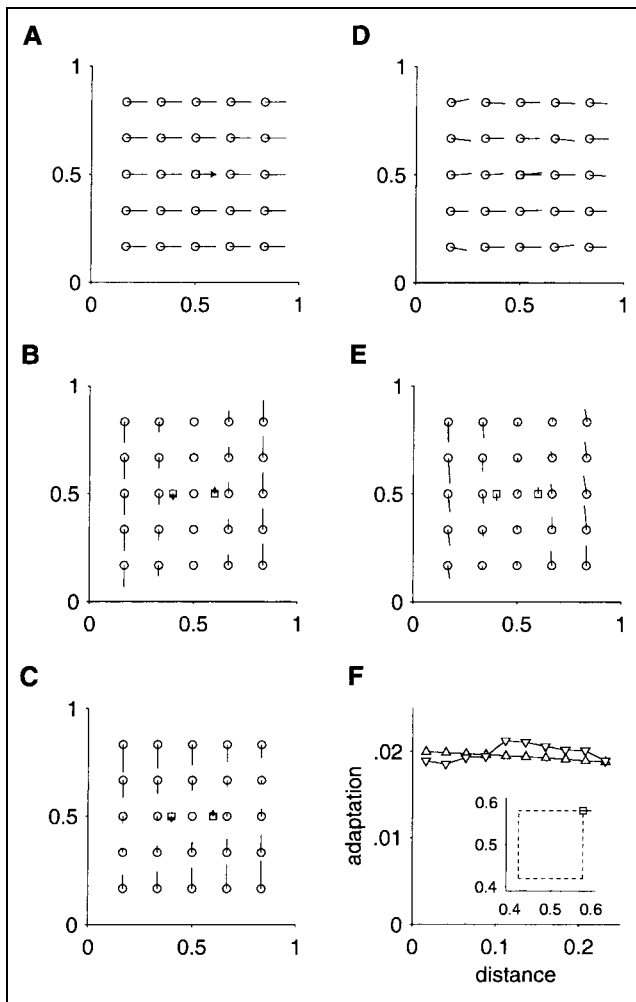


Figure 9. Experiments of Ghahramani et al. (1996). Adapted points are depicted by \square and an arrow. Test points are depicted by \circ . Adaptation is shown by line segments. (A) Adaptation at the central position. (B) Adaptation at two positions. (C) Mapping predicted by the least-square adjustment of the training positions. The line segments have been shortened to improve legibility. (D) Same as (A) in the nonlinear case. (E) Same as (B) in the nonlinear case. (F) Mean change in adaptation as a function of the distance to the exposure point (Δ : linear; ∇ : nonlinear). 2500 points within the restricted workspace (inset) were used.

1989), and it is believed that they share similar computational properties (Pouget & Sejnowski, 1997). In particular, sigmoid functions can be combined to reconstruct tuned functions (Girosi et al., 1995). Here, we adopted a different perspective. We asked what computational mechanism could account for the constrained nature of perceptual-motor adaptations. Clearly, universal approximators are inappropriate for this task because their abilities largely exceed those of the nervous system. Our approach has more in common with basis function models of sensorimotor transformations, in which a single layer of synaptic weights is used to represent the transformations (Salinas & Abbott, 1995; Pouget & Sejnowski, 1994). However, it differs from these models because (1) the monotonic neurons were not combined

with tuned neurons, and (2) the goal was not to reconstruct a tuned activity profile, but a monotonic activity profile. In fact, the model is closely related to the notion of structured representation (Atkeson, 1989), because manipulated variables are readily available in the discharge frequency of input and output neurons. An expected and actually observed property of collective computation in populations of monotonically responding neurons is, thus, the global generalization of learning to nonexperienced situations (Atkeson, 1989). In a recent study (Baraduc et al., 2001), we used this principle to learn a distributed representation of the inverse kinematics of the arm. An appropriate approximation of the desired mapping was obtained following training on a few samples of this mapping. A similar principle was also used by McCandless and Schor (1997) to account for interpolation and extrapolation effects in vertical phoria adaptation (McCandless, Schor, & Maxwell, 1996; Schor et al., 1993). However, in their model, the adapted variable was directly represented in output, and they did not address the case of the distributed representation of this variable. Here, we provide a more general approach to computation between populations of monotonically responding neurons.

The specific component of the model is the sigmoid response function. We therefore need to show that this function is appropriate to encode the dimensions involved in visual-proprioceptive mappings. In the absence of visual landmarks, measuring the direction of a point source in darkness relies on information about eye and head positions (Jeannerod, 1988). The central or peripheral origin of signals related to eye and head posture has not been determined, though the discharge of single neurons in many brain regions is modulated in a monotonic fashion by the static position of the eye and the head (or the gaze) (Bremmer, Pouet, & Hoffman, 1998; Brotchie, Andersen, Snyder, & Goodman, 1995; Andersen, Essick, & Siegel, 1985). The direction of the pointing response toward the point source can be derived from the proprioceptive and efference copy. Neural correlates of static arm posture are found at different levels of somatosensory and motor pathways, and take the form of broad monotonic modulations with variable recruitment thresholds and saturations (Helms Tillery, Soechting, & Ebner, 1996; Gardner & Costanzo, 1981).

The parameter s plays a central role in the model. Its value determines the way the network extrapolates outside the training range, and was chosen to reflect Bedford's conclusions that changes in pointing level off outside the training range. The value of s also influences the strength of the linear constraint on the formation of new mappings. In fact, the choice of s appears as a compromise between extrapolation and linearity: Weak (strong) extrapolation capacities are associated with a weak (strong) linearity constraint. Further data would be necessary to decide on a reasonable s . Our basic

conclusions, however, are not qualitatively altered by the value of s . Still, an open question is the value of s for neurons in the central nervous system. The steepness defines the range of a dimension over which the discharge of a neuron is modulated (i.e., not saturated). Available data from different regions indicate that in general this range encompasses a large portion of the measured range (for eye position: Bremmer et al., 1998; Squatrito & Maioli, 1996; Andersen, Bracewell, Barash, Gnadt, & Fogassi, 1990; for arm position Helms Tillery et al., 1996; Lacquaniti, Guigon, Bianchi, Ferraina, & Caminiti, 1995; Gardner & Costanzo, 1981). However, s is defined by the dimension range (i.e., the range of the recruitment thresholds of neurons), which is difficult to determine (see Materials and Methods).

Bedford's experiments (Bedford, 1989, 1993a, 1993b) can be interpreted in the framework proposed by Redding and Wallace (1997). These authors suggested that prism adaptation is subserved by two mechanisms. "Strategic control" is driven by performance error during exposure and is responsible for the visuomotor reduction of effect (direct effects). "Adaptive spatial alignment" acts to reduce the discordance between expected and achieved effector positions induced by the prism. Adaptation results in a visuomotor negative aftereffect. In the Bedford experiments, the participant's initial pointing response was elaborated by the current internal representation of the visual-proprioceptive correspondence. This response is generally wrong due to the prism-induced distortion and the structure of the internal model. Through trial and error, the participant can discover the correct pointing direction. On the one hand, compensation during exposure is by definition complete from the first trial, because the participant was given feedback about the required response. On the other hand, the discrepancy between the initial and corrected pointing positions is an error signal that slowly drives long-term realignment between the visual and proprioceptive dimensions. The present model is a model of the latter process, which fits the requirements for a mechanism of spatial alignment (Redding & Wallace, 1997): It defines an adjustable parameter-dependent transformation that maintains alignment between spatial dimensions, and can compensate for steady-state discrepancies between the dimensions.

Adaptation was defined as a modified correspondence from a visual to a proprioceptive dimension ($V \rightarrow P$). This description was used to obtain a direct link between experimental and modeling results. In fact, a large component of the adaptation is probably related to changes in the head-hand system, namely, changes in the perceived position of the hand (Bedford, 1993a). Thus, the adaptation should be better conceived as a modified mapping from proprioception to vision, or to a new representation of proprioception. However, the exact nature of adaptation has no influence on the results reported here.

What are the possible extensions of Bedford's theory to multidimensional cases? A first solution is uniform generalization over space. Experimental data do not allow us to dismiss this possibility, but the model is incompatible with this idea. A second solution is LS interpolation among the training pairs, which is not supported by experiments (Ghahramani et al., 1996) or by the model. A third solution is that produced by the model. It is not directly supported by experimental data, but it is not incompatible with the available data. This latter solution provides a reasonable extension of the 1-D case to the 2-D case: (1) uniform generalization in the direction of the remapping and (2) linear interpolation in the direction perpendicular to the remapping. The model suggests that adaptation in the general multidimensional case is not a linear regression. This should not be considered as a limitation of the model as it has been shown experimentally that adaptation is not a linear regression in 2-D (Ghahramani et al., 1996). However, adaptation in 1-D is a particular case where the adaptation is equivalent to a linear regression (Bedford, 1989).

When human subjects are asked to learn associations between stimuli and responses drawn from arbitrary dimensions (function learning), they induce a continuous relation between stimulus and response magnitudes (DeLosh, Bussemeyer, & McDaniel, 1997; Koh & Meyer, 1991; Brehmer, 1974; Carroll, 1963). Further, they display a marked preference for linear relations (DeLosh et al., 1997; Koh & Meyer, 1991; Brehmer, 1974; Carroll, 1963), and response variability is constant across output values (Koh & Meyer, 1991). Function learning, thus, appears to operate much like perceptual-motor learning. The present model constitutes an alternative to rule-based or hybrid rule/exemplar-based approaches to human performance in function learning (DeLosh et al., 1997; Koh & Meyer, 1991). It also offers a neural basis for many high-order cognitive skills (e.g., forecasting, decision making) that require one to have the capacity to discover relations among varying conditions of the environment.

METHODS

General Principle

The principle of the model is the following. A generic mapping between two scalar dimensions x and y is represented by a single-layer linear neural network (Figure 1A). The network is defined by (1) an input layer that encodes x into an activity profile \mathbf{x} ($x \mapsto \mathbf{x} = \mathcal{C}(x) \in \mathbb{R}^N$; Equation 1, below), (2) an output layer that can be decoded to recover y from the activity profile \mathbf{y} ($\mathbf{y} \mapsto y = \mathcal{D}(\mathbf{y}) \in \mathbb{R}$; Equation 2), and (3) a set of synaptic weights \mathbf{W} that establishes a linear correspondence between these dimensions, $\mathbf{y} = \mathbf{W}\mathbf{x}$. The weights can be modified to learn a particular mapping defined by a set of practice pairs. The learning procedure is an error-correction rule (Widrow & Hoff, 1960). Changes in

the synaptic weight between input neuron j and output neuron i are described by

$$\Delta \mathbf{W}_{ij} = \eta x_j (y_i^* - y_i)$$

where y_i^* is the desired output, and η the learning rate. After training, the behavior of the network can be assessed on a set of test pairs. Below we describe encoding and decoding schemes corresponding to families of sigmoid response functions.

Sigmoid Encoding and Decoding Scheme

Consider a population of N neurons. Each neuron i has a mean discharge that varies in a sigmoid fashion with a dimension x in $[a; b]$ ("dimension range") according to

$$x_i = f(x, \lambda_i, s) = \frac{1}{1 + e^{-(x-\lambda_i)/s}} \quad (1)$$

where f is the response function, λ_i the recruitment threshold, and $1/s$ the steepness of f . We assume that the λ_i are uniformly distributed in $[a; b]$. The vector $\mathbf{x} = \mathcal{C}(x) = [x_1 \dots x_N]^T$ defines the encoded representation of x .

The quantity

$$\mathcal{D}(\mathbf{x}) = a + \frac{b-a}{N} \sum_{i=1}^N x_i \quad (2)$$

is an estimator of x (Appendix A). Noise-free decoding is illustrated in Figure 1B (lower inset) for $N = 50$, $s = 5$ (upper inset), and $[a; b] = [-90; 90]$. Errors were confined to the extreme parts of the dimension range. Lower s would reduce these errors. The dimension range should not be confounded with the task range, that is, the actual values of the dimension encountered in a given task, and the physiological range, namely, the maximal realizable values of the dimension (e.g., due to mechanical limitations).

Unlike center-of-mass (Baldi & Heiligenberg, 1988) and population vector (Georgopoulos, Ketter, & Schwartz, 1988) estimators, which apply to the broad tuning case, our estimator is not unbiased because definite decoding errors occur near the extremities of the dimension range. The absence of bias, however, is not an absolute requirement for two reasons. First, systematic biases can be partially avoided by adjusting the dimension range relative to the task range. For a given task range, the larger the dimension range, the smaller the biases. In fact, there is experimental evidence that the dimension range does not coincide with the physiological range. Limb positions that are mechanically impossible can be perceived when effector movement is prevented, but continued changes in muscle afference are induced by artificial means (Craske, 1977). This result suggests that the nervous system possesses receptors whose sensitivity extends outside the physiological range of mechanical parameters. These receptors

could define the dimension range. Second, information derived from proprioception and efferent copy sources is known to be inaccurate (Wann & Ibrahim, 1992). Systematic errors are found when participants point in the dark in a visually specified direction (Bedford, 1989) or toward their unseen hand (Baud-Bovy & Viviani, 1998; van Beers, Sittig, & Denier van der Gon, 1998). The use of unbiased decoding methods would not change qualitatively the results to be reported.

Our encoding/decoding scheme can be extended to the case where the steepness of the response function and the maximal discharge vary among neurons. The decoding method still applies if the distribution of steepness and maximal discharge is the same for each λ_i . This hypothesis can be relieved by using a more efficient decoding method (e.g., LS or maximum likelihood estimation). In a different framework, Pouget, Zhang, Deneve, and Latham (1998) and Zhang, Ginzburg, McNaughton, and Sejnowski (1998) raised the possibility that neural circuits could implement such optimal estimators.

The sigmoid response function was chosen to allow analytical derivations (Appendix A), but, any other S-shaped function (e.g., piecewise linear function) would lead to similar results.

Perceptual-Motor Alignments

A set of experiments on learning new mappings was performed by Bedford (Bedford, 1989, 1993a, 1993b). Participants were required to point to visual targets in the dark. Target location was measured by egocentric angular location V and pointing position by egocentric angle P (Figure 2A). Alignment between V and P (Figure 2B) was distorted by assigning new outputs to a discrete set of inputs using a prism (exposure; Figure 2C). Training positions were in the interval $[-25^\circ; 25^\circ]$, with 0° corresponding to straight ahead. Exposure consisted of a randomized presentation of training pairs. Exposure duration was defined as the number of presentations of each pair (blocks). Adaptation was measured as the change in pointing position $\Delta P = P - P_{\text{pre}}$ for 11 visual positions between -25° and 25° , where P_{pre} is the pointing position before exposure.

Simulations

The network described above was first trained to reproduce the identity mapping $P = V$ (with $x = V$ and $y = P$) over the dimension range $[-90^\circ; 90^\circ]$ (pretraining; Figures 2B and 1B). This interval was chosen to encompass a broad set of visual and pointing locations. Actual values of V and P (in $[-25^\circ; 25^\circ]$) were within a restricted central portion of the range of the response function to avoid decoding errors near the extremities. In this way, patterns of adaptation could not be explained by decoding biases. Pretraining consisted of

100,000 presentations of randomly chosen pairs in the dimension range. Then new mappings were induced as described above (Figure 2C).

Each layer contained two populations of N neurons: One with $s > 0$ (“positive” population) and the other with $-s$ (“negative” population). Output P was decoded from the positive population of the output layer. By symmetry, the negative population could be decoded as well and would provide the same value. Although the two input populations contained the same information, both proved to be necessary to obtain the results reported here. In the same way, although a single-output population is sufficient to recover information from the output layer, the two populations are necessary to convey (input) information to further processing steps.

The purpose of the positive and negative populations is, first of all, related to physiological observations. In general, there should be as many positive and negative neurons related to a given dimension due to the agonist/antagonist organization of postural and motor systems. This characteristic is also necessary for proper functioning of the model. The activity profile of the positive population is a decreasing monotonic pattern that approaches 0 toward rightward positions when encoding leftward positions. Connections arising from the weakly active neurons are little modified during the training period due to the presynaptic term in the learning rule. As a result, there is a nonuniform generalization to untrained values. The activity profile of the negative input population is such that the activity is maximal at positions where activity in the positive population is minimal. Strongly active neurons of the negative population change their outgoing weights during training and compensate for the absence of information in the positive population.

We note that similar results would be obtained using a direct representation instead of a distributed representation of P in the output layer (Pouget & Sejnowski, 1994).

Parameters were $N = 50$, $s = 5$, and $\eta = 0.0005$. All weights were initially set to zero.

APPENDIX

A. Decoding Monotonic Populations

We consider the linear estimator defined by Equation 2. For the sake of simplicity, we consider a normalized dimension X ($[a; b] = [0; 1]$), with normalized recruitment thresholds Λ_i , and steepness S . Equivalence between the normalized and nonnormalized cases is given by

$$\begin{cases} X = (x - a)/(b - a) \\ \Lambda_i = (\lambda_i - a)/(b - a) \\ S = s/(b - a) \end{cases}$$

For a large number of neurons, we can use a continuous approximation and write the estimator as

$$L(X, S) = \int_0^1 f(X, \Lambda, S) d\Lambda$$

One easily shows that

$$L(X, S) = 1 - S \ln \frac{1 + e^{(1-X)/S}}{1 + e^{-X/S}}$$

If we let $S \rightarrow 0$, then $L(X, S) \rightarrow X$. Thus $L(X, S)$ is an unbiased estimator of X for small S .

B. Least-Square Error Learning and Least-Square Approximation

Here we demonstrate that for linear response functions LS error learning in the network is mathematically equivalent to the LS approximation of the training set. There exists a shorter, but less instructive proof for this result.

We use the following notations. Encoding in the input layer is defined by the affine mapping $\mathcal{C}^{\text{in}}: \mathbb{R} \mapsto \mathbb{R}^N, x \mapsto \mathcal{C}^{\text{in}}(x) = x \mathbf{A}^{\text{in}} + \mathbf{B}^{\text{in}}$, with \mathbf{A}^{in} and $\mathbf{B}^{\text{in}} \in \mathbb{R}^N$ (N is the number of neurons). In the same way, encoding in the output layer is defined by the affine mapping $\mathcal{C}^{\text{out}}: \mathbb{R} \mapsto \mathbb{R}^N, x \mapsto \mathcal{C}^{\text{out}}(x) = x \mathbf{A}^{\text{out}} + \mathbf{B}^{\text{out}}$, with \mathbf{A}^{out} and $\mathbf{B}^{\text{out}} \in \mathbb{R}^N$. We assume without loss of generality that $\|\mathbf{A}^{\text{out}}\| = 1$. Decoding the output layer is defined by $\mathcal{D}^{\text{out}}: \mathbb{R}^N \mapsto \mathbb{R}, X \mapsto \mathcal{D}^{\text{out}}(X) = (\mathbf{A}^{\text{out}})^T (X - \mathbf{B}^{\text{out}})$. It can be shown that this decoding scheme provides the maximum likelihood estimate of the encoded parameter for Gaussian and Poisson noise.

The network can be trained to produce a mapping defined by the training set $\{x_t, y_t\}$, ($1 \leq t \leq M, M \geq 2$) using LS error learning, i.e., finding a perceptron $\hat{\mathcal{P}}$ (defined by a matrix \mathbf{W}) that minimizes

$$E(\mathcal{P}) = \sum_t \|\mathcal{C}^{\text{out}}(y_t) - (\mathcal{P} \circ \mathcal{C}^{\text{in}})(x_t)\|^2$$

The actual mapping defined by the network is then the linear mapping $y = (\mathcal{D}^{\text{out}} \circ \hat{\mathcal{P}} \circ \mathcal{C}^{\text{in}})(x)$, which can be written

$$y = (\mathbf{A}^{\text{out}})^T \mathbf{W} \mathbf{A}^{\text{in}} x + (\mathbf{A}^{\text{out}})^T \mathbf{W} \mathbf{B}^{\text{in}} - (\mathbf{A}^{\text{out}})^T \mathbf{B}^{\text{in}}$$

We want to show that this mapping is exactly that defined by linear regression on the training set, that is

$$\begin{cases} (\mathbf{A}^{\text{out}})^T \mathbf{W} \mathbf{A}^{\text{in}} = K\alpha \\ (\mathbf{A}^{\text{out}})^T \mathbf{W} \mathbf{B}^{\text{in}} - (\mathbf{A}^{\text{out}})^T \mathbf{B}^{\text{in}} = K\beta \end{cases} \quad (\text{B.1})$$

with

$$(\alpha, \beta) = \arg \min_{a,b} \sum_t [y_t - (ax_t + b)]^2$$

and K is a constant.

We note \mathbf{w}_i^T the i th row of \mathbf{W} , and we write

$$E(\mathcal{P}) = \sum_i E_i(\mathbf{w}_i)$$

with

$$E_i(\mathbf{w}) = \sum_t (A_i^{\text{out}} y_t + B_i^{\text{out}} - \mathbf{w}^T \mathbf{A}^{\text{in}} x_t - \mathbf{w}^T \mathbf{B}^{\text{in}})^2$$

We assume that $A_i^{\text{out}} \neq 0$. The vector \mathbf{w}_i that minimizes $E_i(\mathbf{w})$ verifies

$$\begin{cases} \mathbf{w}_i^T \mathbf{A}^{\text{in}} = A_i^{\text{out}} \alpha \\ \mathbf{w}_i^T \mathbf{B}^{\text{in}} = B_i^{\text{out}} + A_i^{\text{out}} \beta \end{cases} \quad (\text{B.2})$$

This linear system has a solution in w_i^T iff \mathbf{A}^{in} and \mathbf{B}^{in} are not parallel, meaning that the encoding scheme \mathcal{C}^{in} contains at least two different response functions. It is immediate that Equation B.2 implies that Equation B.1 is satisfied with $K = 1$. Thus, the network actually calculates the linear regression of the training set.

Acknowledgments

We thank R. Brette, Y. Burnod, E. Koechlin, B. Delord, D. Wolpert, and Z. Ghahramani for fruitful discussions, and Suzanne Corkin for revising the English.

Reprint requests should be sent to Emmanuel Guigon, INSERM U483, Université Pierre et Marie Curie, Boîte 23, 9, quai Saint-Bernard, 75005 Paris, France, or via e-mail: guigon@ccr.jussieu.fr.

REFERENCES

Amirikian, B., & Georgopoulos, A. (2000). Directional tuning profiles of motor cortical cells [Erratum in *Neurosci Res* 37(1):83, 2000]. *Neuroscience Research*, 36, 73–79.

Andersen, R., Bracewell, R., Barash, S., Gnadt, J., & Fogassi, L. (1990). Eye position effects on visual, memory, and saccade-related activity in areas LIP and 7a of Macaque. *Journal of Neuroscience*, 10, 1176–1196.

Andersen, R., Essick, G., & Siegel, R. (1985). Encoding of spatial location by posterior parietal neurons. *Science*, 230, 456–458.

Atkeson, C. (1989). Learning arm kinematics and dynamics. *Annual Review of Neuroscience*, 12, 157–183.

Baldi, P., & Heiligenberg, W. (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics*, 59, 313–318.

Baraduc, P., Guigon, E., & Burnod, Y. (2001). Recoding arm position to learn visuomotor transformations. *Cerebral Cortex*, 11, 906–917.

Baud-Bovy, G., & Viviani, P. (1998). Pointing to kinesthetic targets in space. *Journal of Neuroscience*, 18, 1528–1545.

Bedford, F. (1989). Constraints on learning new mappings between perceptual dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 232–248.

Bedford, F. (1993a). Perceptual and cognitive spatial learning. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 517–530.

Bedford, F. (1993b). Perceptual learning. In D. Medin (Ed.), *The psychology of learning and motivation* (vol. 30, pp. 1–60). New York: Academic Press.

Brehmer, B. (1974). Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Decision Processes*, 11, 1–27.

Bremmer, F., Pouget, A., & Hoffmann, K.-P. (1998). Eye position encoding in the macaque posterior parietal cortex. *European Journal of Neuroscience*, 10, 153–160.

Brotchie, P., Andersen, R., Snyder, L., & Goodman, S. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375, 232–235.

Carroll, J. (1963). *Function learning: The learning of continuous functional mappings relating stimulus and response continua (ETS RB 63–26)*. Princeton, NJ: Educational Testing Service.

Craske, B. (1977). Perception of impossible limb positions induced by tendon vibration. *Science*, 196, 71–73.

Dean, J. (1990). Coding proprioceptive information to control movement to a target: Simulation with a simple neural network. *Biological Cybernetics*, 63, 115–120.

DeLosh, E., Busemeyer, J., & McDaniel, M. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 968–986.

Gardner, E., & Costanzo, R. (1981). Properties of kinesthetic neurons in somatosensory cortex of awake monkeys. *Brain Research*, 214, 301–319.

Georgopoulos, A., Kettner, R., & Schwartz, A. (1988). Primate motor cortex and free arm movements to visual targets in 3-dimensional space: II. Coding of the direction of movement by a neuronal population. *Journal of Neuroscience*, 8, 2928–2937.

Ghahramani, Z., Wolpert, D., & Jordan, M. (1996). Generalization to local remappings of the visuomotor coordinate transformation. *Journal of Neuroscience*, 16, 7085–7096.

Girosi, F., Jones, M., & Poggio, T. (1995). Regularization theory and neural networks architectures. *Neural Computation*, 7, 219–269.

Hay, J. (1974). Motor-transformation learning. *Perception*, 3, 487–496.

Held, R. (1965). Plasticity in sensorimotor systems. *Scientific American*, 213, 84–94.

Helms Tillery, S., Soechting, J., & Ebner, T. (1996). Somatosensory cortical activity in relation to arm posture: Nonuniform spatial tuning. *Journal of Neurophysiology*, 76, 2423–2438.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward network are universal approximators. *Neural Networks*, 2, 359–366.

Jeannerod, M. (1988). *The neural and behavioural organization of goal-directed movements*. Oxford: Clarendon Press.

Koh, K., & Meyer, D. (1991). Function learning: Induction of continuous stimulus–response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 811–836.

Lacquaniti, F., Guigon, E., Bianchi, L., Ferraina, S., & Caminiti, R. (1995). Representing spatial information for limb movement: Role of area 5 in the monkey. *Cerebral Cortex*, 5, 391–409.

McCandless, J., & Schor, C. (1997). A neural net model of the adaptation of binocular vertical eye alignment. *Network: Computation in Neural Systems*, 8, 55–70.

McCandless, J., Schor, C., & Maxwell, J. (1996). A cross-coupling model of vertical vergence adaptation.

- IEEE Transactions on Biomedical Engineering*, 43, 24–34.
- Olson, C., & Hanson, S. (1990). Spatial representation of the body. In S. Hanson & C. Olson (Eds.), *Connectionist modeling and brain function* (pp. 193–254). Cambridge: MIT Press.
- Poggio, T., & Girosi, F. (1990). Regularization algorithms for learning that are equivalent to multilayer networks. *Science*, 247, 978–982.
- Pouget, A., & Sejnowski, T. (1994). A neural model of the cortical representation of egocentric distance. *Cerebral Cortex*, 4, 314–329.
- Pouget, A., & Sejnowski, T. (1997). Spatial transformations in the parietal cortex using basis functions. *Journal of Cognitive Neuroscience*, 9, 222–237.
- Pouget, A., Zhang, K., Deneve, S., & Latham, P. (1998). Statistically efficient estimation using population coding. *Neural Computation*, 10, 373–401.
- Redding, G., & Wallace, B. (1997). *Adaptive spatial alignment*. Hillsdale, NJ: Erlbaum.
- Salinas, E., & Abbott, L. (1995). Transfer of coded information from sensory to motor networks. *Journal of Neuroscience*, 15, 6461–6474.
- Schor, C., Gleason, G., Maxwell, J., & Lunn, R. (1993). Spatial aspects of vertical phoria adaptation. *Vision Research*, 33, 73–84.
- Shinn-Cunningham, B., Durlach, N., & Held, R. (1998). Adapting to supernormal auditory localization cues: II. Constraints on adaptation of mean response. *Journal of the Acoustical Society of America*, 103, 3667–3676.
- Snippe, H. (1996). Parameter extraction from population codes: A critical assessment. *Neural Computation*, 8, 511–529.
- Squatrito, S., & Maioli, M. (1996). Gaze field properties of eye position neurones in areas MST and 7a of the macaque monkey. *Visual Neuroscience*, 13, 385–398.
- van Beers, R., Sittig, A., & Denier van der Gon, J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, 122, 367–377.
- Vetter, P., Goodbody, S., & Wolpert, D. (1999). Evidence for an eye-centered spherical representation of the visuomotor map. *Journal of Neurophysiology*, 81, 935–939.
- Wann, J., & Ibrahim, S. (1992). Does limb proprioception drift? *Experimental Brain Research*, 91, 162–166.
- Welch, R. (1978). *Perceptual modification: Adapting to altered sensory environments*. San Diego, CA: Academic Press.
- Welch, R. (1986). Adaptation of space perception. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (vol. 1, pp. 1–45). New York: John Wiley.
- Widrow, B., & Hoff, M. (1960). *Adaptive switching circuits (1)* (Technical Report). Stanford University.
- Zhang, K., Ginzburg, I., McNaughton, B., & Sejnowski, T. (1998). Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology*, 79, 1017–1044.
- Zipser, D., & Andersen, R. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, 331, 679–684.