LETTER **Communicated by Linda Bowns**

Motion-Based Prediction Is Sufficient to Solve the Aperture Problem

Laurent U. Perrinet
Laurent.Perrinet@univ-amu.fr
Guillaume S. Masson
guillaume.masson@univ-amu.fr
Institut de Neurosciences de la Timone, CNRS/Aix-Marseille University 13385 Marseille Cedex 5, France

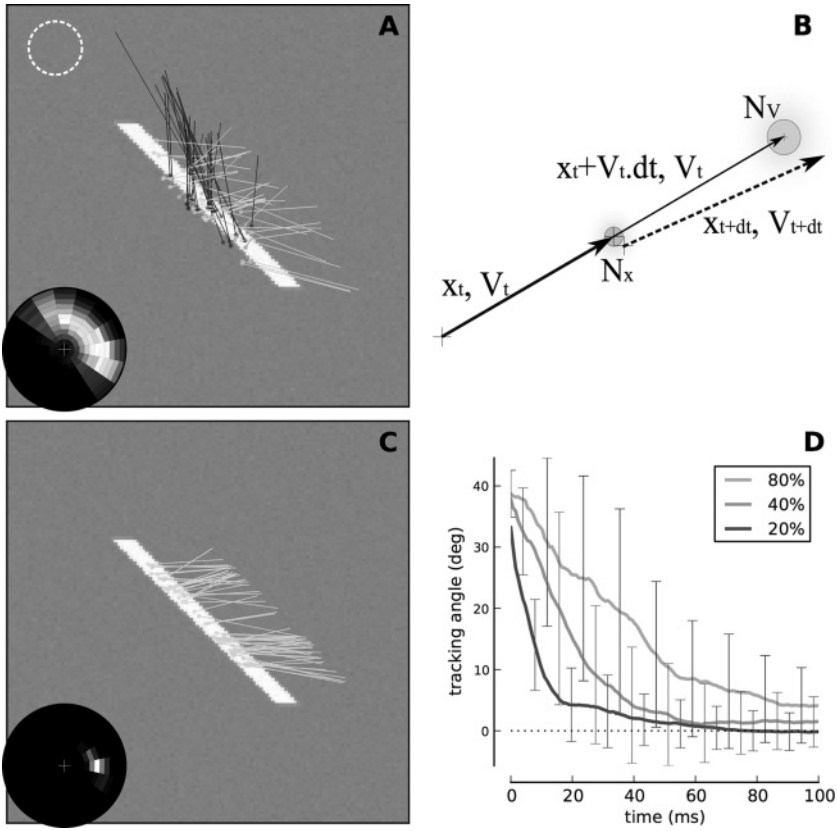
In low-level sensory systems, it is still unclear how the noisy information collected locally by neurons may give rise to a coherent global percept. This is well demonstrated for the detection of motion in the aperture problem: as luminance of an elongated line is symmetrical along its axis, tangential velocity is ambiguous when measured locally. Here, we develop the hypothesis that motion-based predictive coding is sufficient to infer global motion. Our implementation is based on a context-dependent diffusion of a probabilistic representation of motion. We observe in simulations a progressive solution to the aperture problem similar to physiology and behavior. We demonstrate that this solution is the result of two underlying mechanisms. First, we demonstrate the formation of a tracking behavior favoring temporally coherent features independent of their texture. Second, we observe that incoherent features are explained away, while coherent information diffuses progressively to the global scale. Most previous models included ad hoc mechanisms such as end-stopped cells or a selection layer to track specific luminance-based features as necessary conditions to solve the aperture problem. Here, we have proved that motion-based predictive coding, as it is implemented in this functional model, is sufficient to solve the aperture problem. This solution may give insights into the role of prediction underlying a large class of sensory computations.

1 Introduction

1.1 Problem Statement. A central challenge in neuroscience is to explain how local information that is represented in the activity of single neurons can be integrated to enable global and coherent responses at population and behavioral levels. A classic illustration of this problem is given by the early stages of visual motion processing. Visual cortical areas, such as the primary visual cortex (V1) or the mediotemporal (MT) extrastriate area, can extract geometrical structures from luminance changes that are

sensed by large populations of direction- and speed-selective neurons within topographically organized maps (Hildreth & Koch, 1987). However, these cells have access to only the limited portion of the visual space falling inside their classical receptive fields. Therefore, local information is often incomplete and ambiguous, as, for instance, when measuring the motion of a long line that crosses their receptive field. Because of the symmetry along the line's axis, the measure of the tangential component of translation velocity is completely ambiguous, leading to the aperture problem (see Figure 1A). As a consequence, most V1 and MT neurons indicate the slowest element from the family of vectors compatible with the line's translation, that is, the speed perpendicular to the line orientation (Albright, 1984). These neurons are often called component-selective cells and can signal only orthogonal motions of local 1D edges from more complex moving patterns. Integrated in area MT, such local preferences introduce biases in the estimated direction and speed of the translating line. A behavioral consequence is that perceived direction of an elongated tilted line is initially biased toward the motion direction orthogonal to its orientation (Born, Pack, Ponce, & Yi, 2006; Lorenceau, Shiffrar, Wells, & Castet, 1992; Masson & Stone, 2002; Pei, Hsiao, Craig, & Bensmaia, 2010; Wallace, Stone, & Masson, 2005). There are, however, other MT neurons, called pattern-selective cells, that can signal the true translation vector corresponding to such complex visual patterns and, hence drive correct, steady-state behaviors (Movshon, Adelson, Gizzi, & Newsome, 1985; Pack & Born, 2001; Rodman & Albright, 1989). Ultimately, these neurons provide a solution similar to the interception of constraints (IOC) (Adelson & Movshon, 1982; Fennema & Thompson, 1979) by combining the information of multiple component cells (for a recent model, see Bowns, 2011).

The classic view is that these pattern-selective neurons integrate information from a large pool of component cells signaling a wide range of directions, spatial frequencies, speeds, and so on (Rust, Mante, Simoncelli, & Movshon, 2006). However, this two-stage, feedforward model of motion integration is challenged by several recent studies that call for more complex computational mechanisms (see Masson & Ilg, 2010, for reviews). First, there are neurons outside area MT that can solve the aperture problem. For instance, V1 end-stopped cells are sensitive to particular features such as line endings and can therefore signal unambiguous motion at a much smaller spatial scale, providing that the edge falls within their receptive field (Pack, Gartland, & Born, 2004). These neurons could contribute to pattern selectivity in area MT (Tsui, Hunter, Born, & Pack, 2010), but this solution only pushes the problem back to earlier stages of cortical motion processing since one must now explain the emergence of end-stopping cells. Second, all neural solutions to the aperture problem are highly dynamic and build up over dozens of milliseconds after stimulus onset (Pack & Born, 2001; Pack et al., 2004; Pack, Livingstone, Duffy, & Born, 2003; Smith, Majaj, & Movshon, 2010). This can explain why perceived direction of motion gradually



changes over time, shifting from component to pattern translation (Lorenceau et al., 1992; Masson & Stone, 2002; Wallace et al., 2005). Classic feedforward models cannot account for such temporal dynamics and its dependence on several properties of the input such as contrast or bar length (Rust et al., 2006; Tsui et al., 2010). Third, classic computational solutions ignore the fact that any object moving in the visual world at natural speeds will travel across many receptive fields within the retinotopic map. Thus, any single, local receptive field will be stimulated over a period of time that is much less than the time constants reported above for solving the aperture problem (see Masson, Montagnini, & Ilg, 2010). Still, single-neuron solutions for ambiguous motion have been documented so far only with conditions where the entire stimulus is presented within the receptive field (Pack et al., 2004) and with the same geometry (Majaj, Carandini, & Movshon, 2007) over dozens of milliseconds.

Thus, there is an urgent need for more generic computational solutions. We recently proposed that diffusion mechanisms within a cortical map can

solve the aperture problem without the need for complex local mechanisms such as end stopping or pooling across spatiotemporal frequencies (Tlapale, Kornprobst, Masson, & Faugeras, 2011; Tlapale, Masson, & Kornprobst, 2010). This approach is consistent with the role of recurrent connectivity in motion integration (Bayerl & Neumann, 2004) and can simulate the temporal dynamics of motion integration in many different conditions. Moreover, it can reverse the perspective that is dominant in feedforward models where local properties such as end stopping, pattern selectivity, or other types of extraclassic receptive fields phenomena are implemented by built-in, specific neuronal detectors. Instead, these properties can be seen as solutions emerging from the neuronal dynamics of the intricate, recursive contributions of feedforward, feedback, and lateral interactions. A vast theoretical and experimental challenge is therefore to elucidate how diffusion models

Figure 1: (A) The estimation of the motion of an elongated, slanted segment (here moving horizontally to the right) on a limited area (such as the dotted circle) leads to ambiguous velocity measurements compared to physical motion: the aperture problem. Arrows represent the velocity vectors that are most likely detected by a motion energy model; shading indicates direction angle. Due to the limited size of receptive fields in sensory cortical areas (such as shown by the dotted white circle), such a problem is faced by local populations of neurons that visually estimate the motion of objects (see the inset). On a polar representation of possible velocity vectors (the cross in the center corresponds to the null velocity and the outer circle to twice the amplitude of physical speed), we plot the empirical histogram of detected velocity vectors. This representation gives a quantification of the aperture problem in the velocity domain. At the onset of motion detection, information is concentrated along an elongated constraint line (white = high probability, black = zero probability). (B) We use the prior knowledge that in natural scenes, motion as defined by its position and velocity follows smooth trajectories. Quantitatively, it means that velocity is approximately conserved and that the position is transported according to the known velocity. We show here such a transition on position and velocity (respectively, $\vec{x}_t$ and $\vec{V}_t$) from time t to $t + dt$, with the perturbation modeling the smoothness of prediction in position and velocity (respectively, $\mathcal{N}_x$ and $\mathcal{N}_V$). (C) Applying such a prior on a dynamical system detecting motion, we show that motion converges to the physical motion after approximately one spatial period (the line moved by twice its height). The read-out of the system converged to the physical motion: Motion-based prediction is sufficient to resolve the aperture problem (see the inset). (D) As observed at the perceptual level (Castet, Lorenceau, Shiffrar, & Bonnet, 1993; Pei, Hsiao, Craig, & Bensmaia, 2010), size and duration of the tracking angle bias decreased with respect to the height of the line. Height was measured relative to a spatial period (respectively, 60%, 40%, and 20%). Here we show the average tracking angle read-out from the probabilistic representation as a function of time, averaged over 20 trials (error bars show one standard deviation).

can be implemented by realistic populations of neurons dealing with noisy inputs.

The aperture problem in vision must be seen as an instance of the more generic problem of information integration in sensory systems. The aperture problem, as well as the correspondence problem, can be seen as a class of underconstrained inverse problems faced by many different sensory and cognitive systems. Interestingly, recent experimental evidence has pointed out strong similarities in the dynamics of the neural solution for spatiotemporal integration of information in space. For instance, there is a tactile counterpart of the visual aperture problem, and neurons in the monkey somatosensory cortex exhibit similar temporal dynamics to that of area MT neurons (Pei, Hsiao, & Bensmaia, 2008; Pei et al., 2010; Pei, Hsiao, Craig, & Bensmaia, 2011). These recent results point to the need to build a theoretical framework that can unify these generic mechanisms such as robust detection and context-dependent integration and propose a solution that applies to different sensory systems. An obvious candidate is to build association fields that would gather neighboring information and enhance constraints on the response. This is the goal of our study: to provide a theoretical framework using probabilistic inference.

In this letter, we explore the hypothesis that the aperture problem can be solved using predictive coding. We introduce a generic probabilistic framework for motion-based prediction as a specific dynamical spatiotemporal diffusion process on motion representation, as originally proposed by Burgi, Yuille, and Grzywacz (2000). However, we do not perform an approximation of the dynamics of probabilistic distributions using a neural network implementation, as they did. Instead, we develop a method to simulate precise predictions in topographic maps. We test our model against behavioral and neuronal results that are signatures of the key properties of primate visual motion detection and integration. Furthermore, we demonstrate that several properties of low-level motion processing (i.e., feature motion tracking, texture-independent motion, context-dependent motion integration) naturally emerge from predictive coding within a retinotopic map. Finally, we discuss the putative role of prediction in generic neural computations.

1.2 Probabilistic Detection of Motion. First, we define a generic probabilistic framework for studying the aperture problem and its solution. Translation of an object in the planar visual space at a given time is fully given by the probability distribution of its position and velocity, that is, as a distribution of the different degrees of belief among a set of possible velocities. It is usual to define motion probability at any given location. If one particular velocity is certain, its probability becomes 1, while other probabilities are 0. The more the measurement is uncertain (e.g., when increasing noise), the more the distribution of probabilities will be spread around this peak. This type of representation can be successfully used to

solve a large range of problems related to visual motion detection. These problems belong in all generality to optimal detection problems of a signal perturbed by different sources of noise and ambiguity. In particular, the aperture problem is explicitly described by an elongated probability distribution function (PDF) along the constraint defined by the orientation of the line (see Figure 1A inset). This constitutes an ill-posed inverse problem, as different possible velocities may correspond to the physical motion of the line.

In such a framework, Bayesian models make explicit the optimal integration of sensory information with prior information. These models may be decomposed in three stages. First, one defines likelihoods as a measure of belief knowing the sensory data. This likelihood is based on the definition of a generative model. Second, any prior distribution—that is, any information known before observing these data—may be combined to the likelihood distribution to compute a posterior probability using Bayes' rule. The prior defines generic knowledge on the generative model over a set of inputs, such as regularities observed in the statistics of natural images or behaviorally relevant motions. Finally, a decision can be made by optimizing a behavioral cost dependent on this posterior probability. A common choice is the belief that corresponds to the maximum a posteriori probability. The advantage of Bayesian inference compared to other heuristics is that it explicitly states qualitatively and quantitatively all hypotheses (generative models of observation noise and of the prior) that lead to a solution.

1.3 Luminance-Based Detection of Motion. Such a Bayesian scheme can be applied to motion detection using a generative model of the luminance profile in the image. This is first based on the luminance conservation equation. Knowing the velocity $\vec{V}$, we can assume that luminance is approximately conserved along this direction, that is, that after a small lapse dt,

$$I_{t+dt}(\vec{x} + \vec{V} \cdot dt) = I_t(\vec{x}) + \mathcal{N}_I, \tag{1.1}$$

where we define luminance at time t by $I_t(\vec{x})$ as a function of position $\vec{x}$ and $\mathcal{N}_I$ is the observation noise. Using the Laplacian approximation, one can derive the likelihood probability distribution $p(I_t(\vec{x})|\vec{V})$ as a gaussian distribution. In such a representation, precision is finer for a lower variance. Indeed, it is easy to show that the logarithm of $p(I_t(\vec{x})|\vec{V})$ is proportional to the output of correlation-based elementary motion sensors or, equivalently, a motion-energy detector (Adelson & Bergen, 1985). Second, Weiss, Simoncelli, and Adelson (2002) showed that using a prior distribution $p(\vec{V})$ that favors slow speeds, one could explain why the initial perceived direction in the aperture problem is perpendicular to the line. Interestingly, lower-contrast motion results in wider distributions of likelihood and thus posterior $p(\vec{V}|I_t(\vec{x}))$. Therefore, contrast dynamics for a wide variety of

simple motion stimuli is determined by the shape of the probability distribution (i.e., gaussian-like distributions) and the ratio between variances of likelihood and prior distributions, as was validated experimentally on behavioral data (Barthélemy, Perrinet, Castet, & Masson, 2008). With ambiguous inputs, this scheme gives a measure consistent with our formulation of the aperture problem, where probability is distributed along a constraint line defined by the orientation of the line (see Figure 1A, inset).

The generative model explicitly assumes a translational motion $\vec{V}$ over the observation aperture, such as the receptive field of a motion-sensitive cell. Usually a distributed set $\vec{V}_t(\vec{x})$ of motion estimations at time t over fixed positions $\vec{x}$ in the visual field gives a fair approximation of a generic, complex motion that can be represented in a retinotopic map such as V1/MT areas. This provides a field of probabilistic motion measures $p(I_t(\vec{x})|\vec{V}_t(\vec{x}))$. To generate a global read-out from this local information, we may integrate these local probabilities over the whole visual field. Assuming independence of the local information as in Weiss et al. (2002), we model spatiotemporal integration at time T by equation 1.1 and $p(\vec{V}|I_{0:T}) \propto \prod_{\vec{x},0\leq t\leq T} p(I_t(\vec{x})|\vec{V}(\vec{x}))p(\vec{V})$, where we write as $I_{0:t}$ the information on luminance from time 0 to t. Such models of spatiotemporal integration can account for several nonlinear properties of motion integration such as monotonic spatial summation and contrast gain control and are successful in explaining a wide range of neurophysiological and behavioral data. In particular, it is sufficient to explain the dynamics of the solution to the aperture problem if we assume that information from lines and line endings was a priori segmented (Barthélemy et al., 2008). This type of model provides a solution similar to the vector average, and we have previously shown that the hypothesis of an independent sampling cannot account for some nonlinear aspects of motion integration, such as supersaturation of the spatial summation functions, unless some ad hoc mechanisms such as surround inhibition are added (Perrinet & Masson, 2007). In the case of our definition of the aperture problem (see Figure 1A), the information from such Bayesian measurement at every time step will always give the same probability distribution function (described by its mean $\vec{V}_m$ and variance Σ), where $\vec{V}_m$ shows a bias toward the perpendicular of the line (see Figure 1A, inset). The independent integration of such information will therefore necessarily lead to finer precision (the variance becomes Σ/T) but with always the same mean: the aperture problem is not solved.

1.4 Motion-Based Predictive Coding. Failure of the feedforward models in accounting for the dynamics of global motion integration originates from the underlying hypothesis of independence of motion signals in neighboring parts of visual space. The independence hypothesis set out above formally states that the local measurement of global motion is the same everywhere, independent of the position of different motion parts. In fact, the independence hypothesis assumes that if local motion signals were

randomly shuffled in position, they would still yield the same global motion output (e.g., Movshon et al., 1985). As Watamaniuk, McKee, and Grzywacz (1995) showed, this hypothesis is particularly at stake for motion along coherent trajectories: motion as a whole is more than the sum of its parts. A solution used in previous models solving the aperture problem is to add some heuristics, such as a selection process (Nowlan & Sejnowski, 1995; Weiss et al., 2002) or a constraint that motion is relatively smooth away from luminance discontinuities (Tlapale, Masson, et al., 2010). A first assumption is that the retinotopic position of motion is an essential piece of information to be represented. In particular, in order to achieve finely grained predictions, it is essential to consider that the spatial position of motion $\vec{x}$, instead of being a given parameter (classically, a value on a grid), is an additional random variable for representing motion along with $\vec{V}$. Compared to the representation $p(\vec{V}(\vec{x})|I)$ used in previous studies (Burgi et al., 2000; Weiss et al., 2002), the probability distribution $p(\vec{x}, \vec{V}|I)$ more completely describes motion by explicitly representing its spatial position jointly with its velocity. Indeed, it is more generic, as it is possible to represent any distribution $p(\vec{V}(\vec{x})|I)$ with a distribution $p(\vec{x}, \vec{V}|I)$, while the reverse is not true without knowing the spatial distribution of the position of motion $p(\vec{x}|I)$. This introduces an explicit representation of the segmentation of motion in visual space, an essential ingredient in motion-based predictive coding.

Here, we explore the hypothesis that we may take into account most dependence of local motion signals between neighboring times and positions by implementing a predictive dependence of successive measurements of motion along a smooth trajectory. In fact, we know a priori that natural scenes are predictable due to both the rigidity and inertia of physical objects. Due to the projection of their motion in visual space, visual objects preferentially follow smooth trajectories (see Figure 1B). We may implement this constraint into a generative model by using the transport equation on motion itself. This assumes that at time t, during the small lapse dt, motion was translated proportional to its velocity:

$$\vec{x}_{t+dt} = \vec{x}_t + \vec{V}_t \cdot dt + \mathcal{N}_{\vec{x}}, \tag{1.2}$$

$$\vec{V}_{t+dt} = \vec{V}_t + \mathcal{N}_{\vec{V}}, \tag{1.3}$$

where $\mathcal{N}_{\vec{x}}$ and $\mathcal{N}_{\vec{V}}$ are, respectively, position and velocity unbiased noises on the motion's trajectory. In the noiseless case, on the limit when dt tends at zero, this is the auto-advection term in the Navier-Stokes equations and thus implements a fluid prior in the inference of local motion. In fact, it is important to properly tune $\mathcal{N}_{\vec{x}}$ and $\mathcal{N}_{\vec{V}}$ since the variance of these distributions explicitly quantifies the precision of the prediction (see Figure 1B).

We may now use this generative model to integrate motion information. Assuming for simplicity that sensory representation is acquired at discrete,

regularly spaced times, we define integration using a Markov random chain on joint random variables $z_t = \vec{x}_t, \vec{V}_t$:

$$p(z_t|I_{0:t-dt}) = \int_{dz_{t-dt}} p(z_t|z_{t-dt}) \cdot p(z_{t-dt}|I_{0:t-dt}), \tag{1.4}$$

$$p(z_t|I_{0:t}) = p(I_t|z_t) \cdot p(z_t|I_{0:t-dt})/p(I_t|I_{0:t-dt}). \tag{1.5}$$

To implement this recursion, we first compute $p(I_t|z_t)$ from the observation model, equation 1.1. The predictive prior probability $p(z_t|z_{t-dt})$, that is, $p(\vec{x}_t, \vec{V}_t|\vec{x}_{t-dt}, \vec{V}_{t-dt})$, is defined by the generative model defined in equations 1.2 and 1.3. Note that prediction (see equation 1.4) always increases the variance by diffusing information. On the other hand, during estimation (see equation 1.5), coherent data increase the precision of the estimation, while incoherent data increase the variance. This balance between diffusion and reaction will be the most important factor for the convergence of the dynamical system. Overall, these master equations, along with the definition of the prior transition $p(z_t|z_{t-dt})$, define our model as a dynamical system with a simple global architecture but with complex recurrent loops (see Figure 2).

Unfortunately, the dimensionality of the probabilistic representation makes impossible the implementation of realistic simulations of the full dynamical system on classical computer hardware. In fact, even with a moderate quantization of the relevant representation spaces, computing integrals over hidden variables in the filtering and prediction equations (respectively, equations 1.4 and 1.5) leads to a combinatorial explosion of parameters that is intractable with the limited memory of current sequential computers. Alternatively, if we assume that all probability distributions are gaussian, this formulation is equivalent to Kalman filtering on joint variables. This type of implementation may be achieved using, for instance, a neuromorphic approximation of the above equations (Burgi et al., 2000). Indeed, one may assume that master equations are implemented by a finely tuned network of lateral and feedback interactions. One advantage of this recursive definition in the master equations is that it gives a simple framework for implementing association fields. However, this implementation has the consequence of blurring predictions. We explore another route using the condensation algorithm (Isard & Blake, 1998), which surpasses the above approximation. In addition, it allows us to explore the role of prediction in solving the aperture problem on a more generic level.

2 Model and Methods

2.1 Particle Filtering. Master equations can be approximated using sequential Monte Carlo (SMC). This method (also known as particle filters)

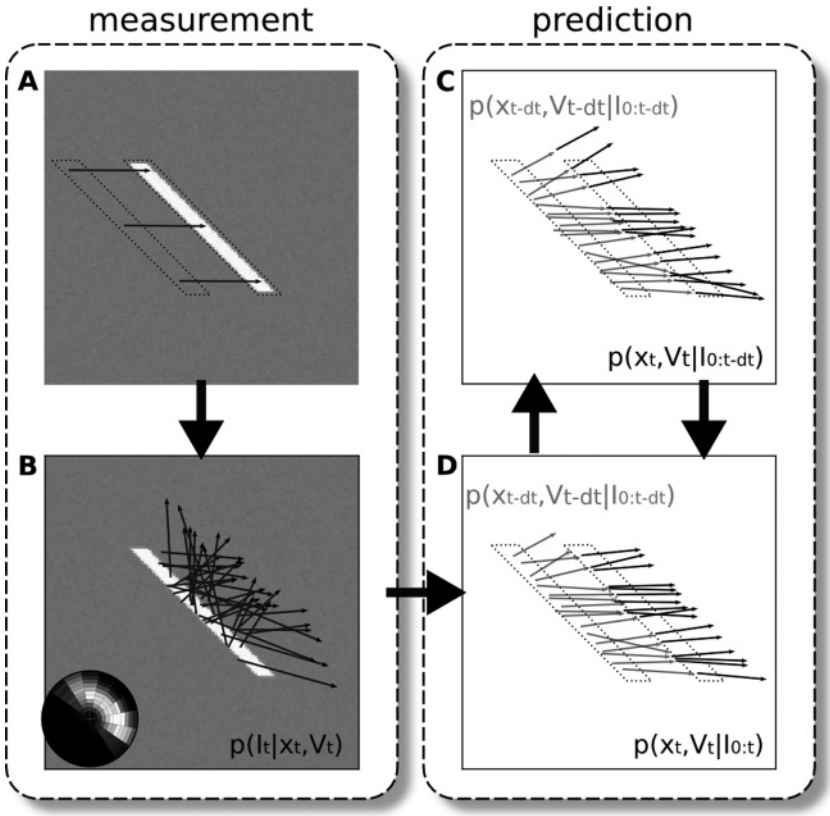


Figure 2: Architecture of the model. The model is constituted by a classical measurement stage and a predictive coding layer. The measurement stage consists of (A) inferring from two consecutive frames of the input flow (B) a likelihood distribution of motion. This layer interacts with the predictive layer, which consists of (C) a prediction stage that infers from the current estimate and the transition prior to the upcoming state estimate and (D) an estimation stage that merges the current prediction of motion with the likelihood measured at the same instant in the previous layer (B).

is a special version of importance sampling in which PDFs are represented by weighted samples. Here, we represent joint variables of motion as a set of four-dimensional vectors merging position $\vec{x} = (x, y)$ and velocity $\vec{V} = (u, v)$. Using sampling, any distribution $p(\vec{x}, \vec{V}|I)$ may be approximated by a set of weighted samples, or "particles," $\pi^{1:N} = \{\pi^i\}_{i\in 1:N} = \{(x^i, y^i, u^i, v^i)\}_{i\in 1:N}$, along with weights $w^{1:N} = \{w^i\}_{i\in 1:N}$. Weights are positive

($\forall i, w^i \geq 0$) and normalized ($\sum_{i=1}^{N} w^i = 1$). By definition, $p(\vec{x}, \vec{V}|I) \approx \hat{p}(\vec{x}, \vec{V}|I)$ with

$$\hat{p}(\vec{x}, \vec{V}|I) = \sum_{i \in 1:N} w^i \cdot \delta(\vec{x} - (x^i, y^i), \vec{V} - (u^i, v^i)), \tag{2.1}$$

where δ is the Dirac measure. There are many different sampling solutions to one given PDF. Prototypical solutions are either a uniform sampling of position and velocity spaces with weights proportional to $p(\vec{x}, \vec{V}|I)$ or the sampling corresponding to uniform weights with a density of samples proportional to the PDF. Compared to other approximations, such as the Laplacian approximation of the PDF by a gaussian, this representation has the advantage of allowing the representation of arbitrary distributions, such as the sparse or multimodal distributions that are often encountered with natural scenes.

This weighted sample representation makes the implementation of equations 1.4 and 1.5 tractable on a sequential computer. To initialize the algorithm, we set particles $\pi_{t=0}^{1:N}$ to random values with uniform weights. Then the following two steps are repeated in order to recursively compute particles $\pi_t^{1:N}$. This set represents $\hat{p}(\vec{x}_t, \vec{V}_t|I_{0:t})$, while particles $\pi_{t-dt}^{1:N}$ represent $\hat{p}(\vec{x}_{t-dt}, \vec{V}_{t-dt}|I_{0:t-dt})$. First, the prediction equation implemented by equation 1.4, which uses the prior predictive knowledge on the smoothness of the trajectory, may be implemented to each particle by a deterministic shift followed by a diffusion such as defined in the generative model (see equations 1.2 and 1.3). The noise $\mathcal{N}_\pi = (\mathcal{N}_{\vec{x}}, \mathcal{N}_{\vec{V}})$ is described here as a four-dimensional centered and decorrelated gaussian. This intermediate set of particles represents an approximation of $p(\vec{x}_t, \vec{V}_t|I_{0:t-dt})$.

Second, we update measures as recorded by the observation likelihood distribution $\hat{p}(I_t|\vec{x}_t, \vec{V}_t)$ computed from the input sensory flow. As in the SMC algorithm, we apply equation 1.5 using the sampling approximation by updating the weights of the particles: $\forall i$,

$$w_t^i = 1/Z \cdot w_{t-dt}^i \cdot p(I_t|\pi_t^i), \tag{2.2}$$

where the scalar Z ensures normalization of the weights ($\sum_{i=1}^{N} w_t^i = 1$). Likelihood $p(I_t|\pi_t^i)$ is computed using equation 1.1 thanks to a standard method, which we describe below. Finally, a usual numerical problem with SMC is the presence of particles with small weights due to sample impoverishment. We use a classical resampling method that results in eliminating particles that assign lower weights (as detected by a threshold relative to the average weight) and duplicating particles that assign the largest weights. In analogy with an homeostatic transform, it has the property of distributing resources while not changing the representation (Perrinet, 2010). In

summary, this formulation is similar to the processing steps in the condensation algorithm which was used in another context for tracking moving shapes (Isard & Blake, 1998), and we apply it here to motion-based prediction as defined in equations 1.2 and 1.3. Note that although our probabilistic approach is exactly similar to that of Burgi et al. (2000) at the computational and algorithmic levels, the actual implementation is completely different. The approach of Burgi et al. (2000) seeks to achieve an implementation close to neural networks. We will see that using the particle filtering approach, though a priori less neuromorphic, allows achieving higher precision in motion-based prediction, which is essential in observing the emergence of complex behavior characteristics of neural computations.

An advantage of importance sampling is that it allows easily computing moments of the distribution. This is particularly useful to define different read-out mechanisms in order to compare the output of our model with biological data. For instance, we can compute the read-out for tracking eye movements as the best estimator (i.e., the conditional mean) using the approximation of $p(\vec{x}, \vec{V}|I)$:

$$\langle \vec{V}|I \rangle = \int p(\vec{x}, \vec{V}|I) \cdot \vec{V} \cdot d\vec{V} \cdot d\vec{x} \approx \begin{pmatrix} \sum_{i=1}^{N} w^i \cdot u^i \\ \sum_{i=1}^{N} w^i \cdot v^i \end{pmatrix}. \tag{2.3}$$

Furthermore, by restricting the integration to a subpopulation of neurons, we can also compare a model output with single-neuron selectivity and thus test how neuronal properties such as contrast gain control or center-surround interactions could emerge from such predictive coding.

2.2 Numerical Simulations. The SMC algorithm itself is controlled by only two parameters. The first is the number of particles N, which tunes the algorithmic complexity of the representation. In general, N should be large enough; an order of magnitude of $N \approx 2^{10}$ was always sufficient in our simulations. In the experimental settings that we have defined here (moving dots or lines), the complexity of the scene is controlled and low. Control experiments have tested the behavior for different numbers of particles (from 2^5 to 2^{16}) and have shown, except for N smaller than 100, that results were always similar. However, we held N to this quite high value to keep the generality of the results for further extensions of the model. The other parameter is the threshold for which particles are resampled. We found that this parameter had little qualitative influence providing that its value is large enough to avoid staying in local minima. Typically a resampling threshold of 20% was sufficient.

Once the parameters of the SMC were fixed, the only free parameters of the system were the variances used to define the likelihood and the noise model $\mathcal{N}_\pi$. Likelihood of sensory motion was computed using equation 1.1

using the same method as Weiss et al. (2002). We defined space and time as the regular grid on the toroidal space to avoid border effects. Next, visual inputs were 128×128 grayscale images on 256 frames. All dimensions were set in arbitrary units, and we defined speed such that $V = 1$ corresponds in toroidal space to the velocity of one spatial period within one temporal period that we defined arbitrarily to 100 ms biological time. Raw images were preprocessed (whitening, normalization), and we computed at each processing step the likelihood locally at each point of the particle set. This computation was dependent only on image contrast and the width of the receptive field over which likelihood was integrated. We tested different parameter values that resulted in different motion direction or spatiotemporal resolution selectivities. For instance, a larger receptive field size gave a better estimate of velocity but poorer precision for position, and vice versa. Therefore, we set the receptive fields size to a value yielding a good trade-off between precision and locality (i.e., 5% of the image's width in our simulations). Similarly, the contrast of likelihood was tuned to match the average noise value in the set of images. We also controlled that using a prior favoring slow speeds had little qualitative influence on our results, and we used a flat prior on speeds throughout this letter. Once fixed, these two values were kept constant across all simulations. Note that the individual measurements of the likelihood may represent multimodal densities if the corresponding individual motions are more than an order of the receptive field's size (as when tracking multiple dots). However, such measurements may be perturbed if individual motions are superimposed on a receptive field. Such a generative model of the input may be accounted for by using a gaussian mixture model (Isard & Blake, 1998). The types of stimuli we are considering are always well described by a unimodal distribution, and we here restrict ourselves to this simple formulation.

All simulations were performed using Python with modules Numpy (Oliphant, 2007) and Scipy (respectively, versions 2.6, 1.5.1, and 0.8.0) on a cluster of Linux nodes. Visualization was performed using Matplotlib (Hunter, 2007). (All source code is available on request from L.U.P.)

3 Results

3.1 Prediction Is Sufficient to Solve the Aperture Problem. Similar to classic studies on a biological solution to the aperture problem, we first used as input the image of a horizontally moving diagonal bar. The initial representation shows a bias toward the perpendicular of the line, as previously found with neuronal (Pack & Born, 2001; Pei et al., 2010), behavioral and perceptual responses (Born et al., 2006; Lorenceau et al., 1992; Masson & Stone, 2002; see Figure 1A). Moreover, the global motion estimation represented by the probability density function converges quickly to the physical motion in terms of both retinotopic position and velocity (see Figure 1C). Changing the length of the line did not qualitatively change the dynamics

but rather proportionally scaled the time it takes to the system for converging to the physical solution (Born et al., 2006; Castet, Lorenceau, Shiffrar, & Bonnet, 1993; see Figure 1D). This result demonstrates that motion-based prediction is sufficient to resolve the aperture problem.

Interestingly, results show that line endings are preferentially tracked (see section 3.3). In fact, the system responds optimally to predictable features, and thus it can correctly detect line endings motion with a probability that is higher than observed for any points located at, say, the middle of the line segment. Moreover, it was shown behaviorally that when blurring the stimulus's line endings, motion representation still converges toward the physical motion, albeit with a slower dynamics (Wallace et al., 2005). This is another key signature that we successfully replicated in our model. This shows that end-stopped cells (or, more generally, local 2D motion detectors, Wilson, Ferrera, & Yo, 1992) are not necessary to solve the aperture problem. On the contrary, a reliable 2D tracking motion system appears to be rather no more than the consequence of cells tuned to predictive trajectories.

This result has some generic consequences that were not described in previous models such as Bayerl and Neumann (2007) and Burgi et al. (2000). First, the emergence of line-ending detectors is caused by the fact that the model filters coherent motion trajectories. This property emerges as line endings follow a coherent trajectory, but this property is not limited to line endings. As a consequence, the most salient difference is that "interesting" features are defined not by a property of the luminance profile but rather by the coherence of their motion's trajectory. Such a distinction is important with regard to biological experiments. Indeed, at the behavioral level, Watamaniuk et al. (1995) have shown that the sequential detection of line endings is not sufficient to explain at a global level the change of behavior when an object moves on a coherent trajectory. Second, at the physiological level, Pack et al. (2003) have shown in the macaque monkey different phases in the dynamics of MT neurons tuned for line endings.This suggests that the selective response to line endings is a consequence of the presentation of a coherent trajectory.

3.2 Emergence of Texture-Independent Motion Trackers. To further understand these mechanisms, we tested the response of the dynamical system to a coherently moving dot. This was defined as a gaussian blob of luminance. Its center moved with a constant translational velocity. For a wide range of parameters, we found that the particles representing the distribution of motion quickly concentrate on the dot's center, while their velocity converged to the true physical velocity. Thanks to the additional information given by the predictive information, this convergence is much quicker than what would be obtained by simply integrating temporally the raw inputs. Moreover, the response of the system is qualitatively different from what is expected in the absence of prediction. In fact, if the

dot's motion is coherent with the predictive generative model, information is either amplified or reduced, resulting in progressively more and more binary responses as time progresses. This behavior is the consequence of the autoreferential formulation of our motion detection scheme. Indeed, precision of motion estimation is modulated by a prediction that is itself estimated using motion. We therefore see the emergence of a basic tracking behavior where the dot's trajectory is captured by the system.

We explored the effects of some key parameters on the tracking behavior of the model. First, when progressively adding uniform gaussian white noise to the stimulus, we found that convergence time to veridical tracking increased with respect to the level of noise. Then, at a certain level of noise, error bias in the prediction becomes larger than required for the balance in tracking amplification and dots are rapidly lost. This can define a tracking sensitivity threshold that can be characterized by plotting the contrast response function of our system (see Figure 3). Second, we varied the precision of prediction. It is quantitatively defined by the inverse variance of the noise present in the generative model (see equations 1.2 and 1.3). We observed that the convergence speed of tracking grew proportionally with this parameter. For very low precision values, the tracking behavior is lost. Moreover, we observed that increasing this prediction's precision above a certain threshold leads to the detection of false positives: An initial movement may be predicted in a false trajectory but is not discarded by sensory data. In fact, this is due to the high positive feedback generated by the high precision assigned to the prediction. In summary, varying both parameters, that is, external and internal variability, we can identify three distinct regimes in this state-space: an area of correct tracking (see Figure 3, TT), an area where there is no tracking due to low precision or high noise (see Figure 3, NT), and an area of false tracking (see Figure 3, FT). These three regimes fully characterize the emergence of the tracking behavior of the dynamical system implementing motion-based prediction.

We then studied how such tracking behavior is independent of the luminance profile of the object being tracked. To achieve that, we tested our system with the same dot but whose envelope was multiplied by a random white noise texture. When this texture consists of a static grating, we obtain one instance of second-order motion (see Lu & Sperling, 2001, for a review). Although the convergence was longer and more variable, tracking was still observed in a robust fashion, and the envelope's motion was ultimately retrieved. This property is due to the fact that in the generative model, we define the prediction as based on both the position and trajectory of motion, independent of the local geometry of image features. This is different from motion detection models, which try to track a particular luminance feature (Lu & Sperling, 2001; Wilson et al., 1992). As a consequence, this dynamical system will have a preference for objects conserving their motion along a trajectory, independent of their texture. Such invariance is usually obtained by introducing, and tuning, a well-known static nonlinear

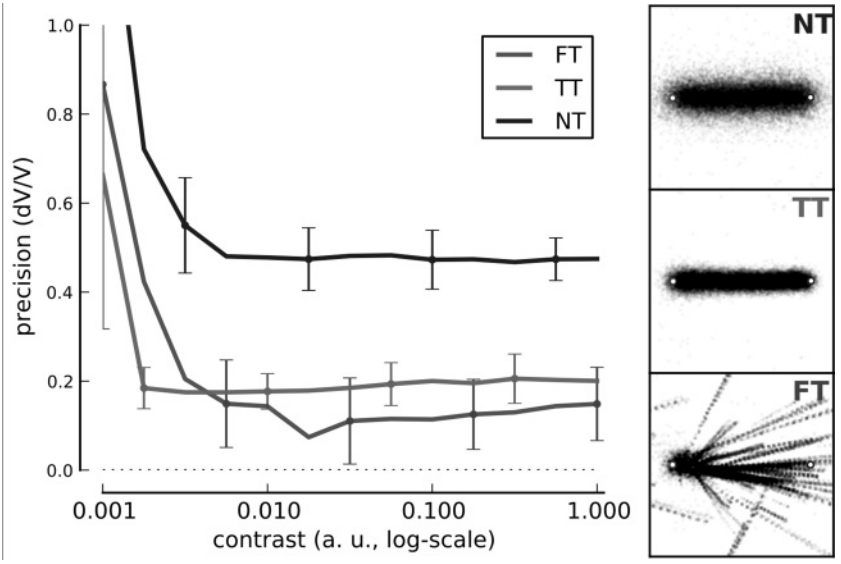
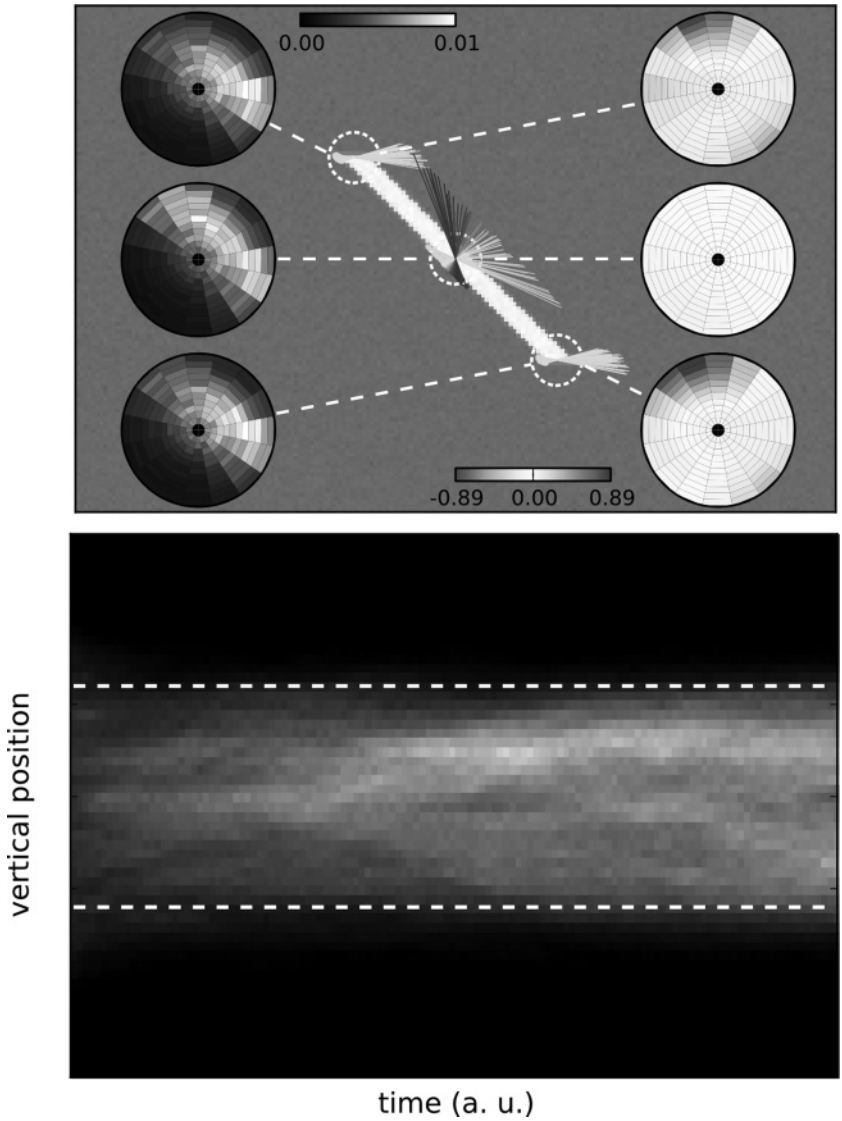


Figure 3: To explore the state-space of the dynamical system, we simulated motion-based prediction for a simple small dot (size 2.5% of a spatial period) moving horizontally from the left to the right of the screen. We tested different levels of sensory noise with respect to different levels of internal noise, that is, to different values of the strength of prediction. (Right) Results show the emergence of different states for different prediction precisions: a regime when prediction is weak and shows high tracking error and variability (no tracking, NT), a phase for intermediate values of prediction strength (as in Figure 1) exhibiting a low tracking error and low variability in the tracking phase (true tracking, TT), and a phase corresponding to higher precisions with relatively efficient mean detection but high variability (false tracking, FT). We give three representative examples of the emerging states at one contrast level ($C = 0.1$) with starting and ending points and, respectively, NT, TT, and FT, by showing inferred trajectories for each trial. (Left) We define tracking error as the ratio between detected speed and target speed and plot it with respect to the stimulus contrast as given by the inverse of sensory noise. Error bars give the variability in tracking error as averaged over 20 trials. As prediction strength increases, there is a transition from smooth contrast response function (NT) to more binary responses (TT and FT).

computation such as divisive normalization (Rust et al., 2006; Simoncelli & Heeger, 1998).

3.3 Role of Context for Solving the Aperture Problem. In order to better understand how the different parts of the line interact over time,



we finally investigated modulation of neighboring motions in the aperture problem. In fact, this also corresponds to the case of the diagonal line. At the initial time step, motion position information is spread preferentially along the edge of the line and represents motion ambiguity with a speed probability distributed along the constraint line (see Figure 1A, inset). In particular, trajectories are inferred on different trajectories preferentially on

the points of the line but with directions that are initially ambiguous due to the aperture problem. These different trajectories evolve independently but are ultimately in competition. To understand the underlying mechanism, we first focus on a single step of the algorithm at three independent key positions of the stimulus: the two edges and the center. Compared with the case without prediction, we show that prediction induces a contextual modulation of the response to different trajectories, such as explaining away trajectories that fall off the line (see Figure 4, top). This modulation acts on a large scale as a gain-control mechanism, which is reminiscent of what is observed in center-surround stimulation.

We can then analyze in greater detail the dynamics of motion distributions for the aperture problem. From the initial step, unambiguous line-ending information spreads progressively towards the rest of the line, progressively explaining away motion signals that are inconsistent with the target speed (see Figure 4, bottom). In fact, from the formulation of prediction in the master equation, probability at a given point reflects the accumulated evidence of each trajectory leading to that point as it is computed by the predictive prior. Combined with likelihood measurements, incoherent trajectories will be progressively explained away as they fall off the line. Such gradual diffusion of information between nearby locations explains the role of line length already documented in Figure 1D, as well as why information takes time to diffuse at the global scale of the stimulus,

---

Figure 4: (Top) Prediction implements a competition between different trajectories. Here, we focus on one step of the algorithm by testing different trajectories at three key positions of the segment stimulus: the two edges and the center (dashed circles). Compared to the pure sensory velocity likelihood (left insets in grayscale), prediction modulates response as shown by the velocity vectors (direction coded as in Figure 1) and by the ratio of velocity probabilities (log ratio in bits, right insets). There is no change for the middle of the segment (light gray tone), but trajectories that are predicted out of the line are "explained away" while others may be amplified (see the color bar). Notice the asymmetry between both edges—the upper edge carrying a suppressive predictive information while the bottom edge diffuses coherent motion. (Bottom) Finally, the aperture problem is solved due to the repeated application of this spatiotemporal contextual information modulation. To highlight the anisotropic diffusion of information over the rest of the line, we plot as a function of time (horizontal axis) the histogram of the detected motion marginalized over horizontal positions (vertical axis), while detected direction of velocity is given by the distribution of tones. Initial tones (left-most column) correspond to the direction perpendicular to the diagonal, while a lighter tone represents a disambiguated motion to the right (as in Figure 1). The plot shows that motion is disambiguated by progressively explaining away incoherent motion. Note the asymmetry in the propagation of coherent information.

as is reported at the physiological level (Pack et al., 2003). In summary, contrary to other models consisting of a selection stage, the system selects coherent features in an autonomous and progressive manner based on the coherence of all their possible trajectories. This ultimately explains why in the aperture problem, information diffuses in the system from line endings to the rest of the segment to ultimately resolve the correct physical motion.

A counterintuitive result is that the leading bottom line ending is less informative that the trailing upper line ending. This was already evident from the asymmetry revealed in Figure 4 (top), which makes explicit that motion-based prediction will have a different effect on both line endings. Indeed, in the leading line ending, most information is diffused to the rest of the line and is not explained away. On the contrary, for the trailing line ending, the diffusion of information is more constrained, as any motion hypothesized to be going upward would soon be explained away from motion-based prediction, as it would fall off the line. This asymmetry is clearly observable in Figure 4 (bottom) as the ambiguous information (distributed on the whole line at the left-most initial time) is progressively resolved by the diffusion of the information originating from the trailing line ending. Unfortunately, the experiments using blurring of the line performed by Wallace et al. (2005) were preformed symmetrically, that is, similar for both edges. We thus predict that blurring the trailing line ending should lead to a greater bias angle as blurring the leading line ending only.

4 Discussion

Our computational model shows that motion-based prediction is sufficient to solve the aperture problem as well as other motion integration phenomena. The aperture problem instantiated with slanted lines in visual space helps to capture several generic computations that are often considered as essential features of any sensory areas. We have shown that predictive coding through diffusion is sufficient to explain the emergence of local 2D motion detectors but also texture-independent motion grabbers. It can also implement context-dependent competition between local motion signals. All of these computations are emerging properties from the dynamics of the system. This view is opposite to the classical assumptions that these mechanisms are implemented by specific, separated mechanisms (Grossberg, Mingolla, & Viswanathan, 2001; Lu & Sperling, 2001; Tsui et al., 2010; Wilson et al., 1992). Instead, we demonstrate here that all of these properties must be seen as the mere consequence of a simple, unifying computational principle. By implementing a predictive field, motion information is anisotropically propagated as modulated by sensory, local estimations such that motion representation dynamically diffuses from a local to a global scale. This model offers a simplification of our original model (Tlapale, Masson, et al., 2010).

4.1 Relation to Other Models. In fact, we can take advantage of the work from Tlapale, Kornprobst, Bouecke, Neumann, and Masson (2010) to compare our model with Tlapale, Masson, et al. (2010) and a large range of models in the community. This study compared the results obtained from different modeling approaches on the same aperture problem and used their model as a reference point, but with two main differences. First, it does not try to make a neuromorphic approach except the fact that (to respect the definition of the aperture problem) information is grabbed locally and propagated on a neighborhood. Moreover, in our model, information is represented explicitly by probabilities, and we make no assumption on how it is represented in the neural activity as this would introduce an unnecessary hypothesis regarding our objective. Second, motion-based prediction defines an anisotropic, context-dependent direction of propagation, while most previous models were using an isotropic diffusion dependent on some feature characteristics (like gating the diffusion by luminance). However, our model explicitly uses the selectivity brought by the anisotropic diffusion. As a consequence, it needs less tuning of the parameters of the diffusion mechanisms, a common problem in the latter type of models. A further advantage of our approach is that it does not contradict previous models. Rather, motion-based prediction seems to be a promising approach to implement in neuromorphic models.

In particular, several parts of our model are similar to previous models of motion detection, although its implementation is radically novel. First, it inherits properties of functional models such as the probabilistic formulation of Weiss et al. (2002) but with simpler hypotheses. For instance, we do not need a prior distribution favoring slow speeds or some selective process needed to preprocess the data (Barthélemy et al., 2008; Weiss et al., 2002). Our model uses a simple Markov chain formulation that has been used for spatial luminance-based prediction or shape tracking with SMC in the Condensation algorithm (Isard & Blake, 1998), but this was to our knowledge not applied to an explicit definition of motion-based prediction. Note that the model presented in Bayerl and Neumann (2007) includes an anisotropic diffusion based on motion-based prediction but that this study was using a neural approximation of the kind that Burgi et al. (2000) used. However, they did not study in particular the role of prediction in the progressive resolution of the aperture problem and its characteristic signature compared to biological data. The application of their fast implementation to our model appears to be a promising perspective. Ultimately our model also gives a more formal description of the dynamical Bayesian model that we originally suggested to implement dynamical inference solution for motion integration (Bogadhi, Montagnini, Mamassian, Perrinet, & Masson, 2011; Montagnini, Mamassian, Perrinet, Castet, & Masson, 2007).

Moreover, when compared to other models designed for understanding visual motion detection (Bayerl & Neumann, 2004; Grossberg et al., 2001; Wilson et al., 1992), our approach is more parsimonious as we do not need

to explicitly model specialized edge detectors. On the contrary, we show that these local feature detectors must be seen as emerging properties from a subset of coherent-motion detectors. Nevertheless, this emergence needs a finely scaled prediction, as we have shown that these properties depend on the precision of prediction. Our computational implementation using SMC could reach higher precision levels compared to the earlier predictive model proposed by Burgi et al. (2000). We could therefore explore a range of parameters and stimuli (such as the aperture problem) that is radically different from the original study. Moreover, some nonlinear behaviors observed in our model are similar to other signatures of linear and nonlinear models such as the cascade model from Rust et al. (2006) or mesoscopic models (Bayerl & Neumann, 2004, 2007; Tlapale, Masson, et al., 2010). However, these last models are specifically tuned by assembling complex and precise knowledge from the dynamical behavior of neurons and their interactions to fit the results that were obtained neurophysiologically. In our model, though, these properties emerge from the interactions in the probabilistic model.

4.2 Toward a Neural Implementation. More generally, this probabilistic and dynamical approach unveils how complex neural mechanisms observed at population levels (or from their read-outs) may be explained by the interactions of local dynamical rules. As mentioned above, both visual (Pack & Born, 2001; Pack et al., 2003, 2004; Smith et al., 2010) and somatosensory (Pei et al., 2010) systems exhibit similar neuronal dynamics when solving the aperture problem or other sensory integration tasks in space and time. This suggests that different sensory cortices might use similar computational principles for integrating sensory inflow into a coherent, nonambiguous representation the motion of objects. By avoiding mechanisms such as neuronal selectivities for some specific local features, our approach offers a more generic framework. It also allows seeking simple, low-level mechanisms underlying complex visual behavior and their dynamics as observed, for instance, with reflexive tracking eye movements (see Masson & Perrinet, 2012, for a review). Finally, we propose that distributions of neural activity on cortical maps act as probabilistic representations of motion over the whole sensory space. This suggests that, for instance, in cortical areas V1 and MT, all probable solutions are initially superposed. This is coherent with the dynamics of the population of MT neurons when solving the aperture problem or computing plaid pattern motion (Pack & Born, 2001; Pack et al., 2004; Smith et al., 2010). Simple decision rules can be applied to these maps to trigger different behaviors such as saccadic and smooth pursuit eye movements as well as perceptual judgments of motion such as direction and speed. Then the temporal dynamics of these behavioral responses can be explained by the dynamics of predictive coding at a sensory stage (Bogadhi et al., 2011).

This work provides new insights for neuroscience and computational paradigms. In fact, biological vision still outperforms any artificial system for simple tasks such as motion segmentation. Our simple model is validated based on neurophysiological and behavioral data and gives several perspectives for its application to image processing. In the future, our model will provide interesting perspectives for exploring novel probabilistic and contextual interactions thanks to the use of neuromorphic implementations. Indeed, it is impossible in practice today to implement the full system on classical von Neumann architectures due to the size of the memory required to implement such complex association fields. However, as we saw above, the probabilistic representation of motion has a natural representation in a neural architecture, where many simple processors are densely connected. Thus, this model is structurally compatible with generic neural architectures and is a candidate for functional implementation on wafer-like hardware. Such recent innovative computing architectures enable constructing specialized neuromorphic systems, allowing new possibilities thanks to their massive parallelism (Brüderle et al., 2011). In return, this approach will allow us to implement models simulating complex association fields. Studying novel computational paradigms in such systems will help extend our understanding of neural computation.

Acknowledgments

This work is supported by EC IP project FP6-015879, "FACETS," and FP7-269921, "BrainScaleS." Code to reproduce figures and supplementary material is available online on L.U.P.'s Web site at http://invibe.net/LaurentPerrinet/Publications/Perrinet12pred.

References

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of Optical Society of America, A*, *2*(2), 284–299.

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, *300*(5892), 523–525.

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*, 1106–1130.

Barthélemy, F. V., Perrinet, L. U., Castet, E., & Masson, G. S. (2008). Dynamics of distributed 1D and 2D motion representations for short-latency ocular following. *Vision Research*, *48*(4), 501–522.

Bayerl, P., & Neumann, H. (2004). Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, *16*, 2041–2066.

Bayerl, P., & Neumann, H. (2007). A fast biologically inspired algorithm for recurrent motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(2), 246–260.

Bogadhi, A. R., Montagnini, A., Mamassian, P., Perrinet, L. U., & Masson, G. S. (2011). Pursuing motion illusions: A realistic oculomotor framework for Bayesian inference. *Vision Research*, *51*, 867–880.

Born, R. T., Pack, C. C., Ponce, C. R., & Yi, S. (2006). Temporal evolution of 2-dimensional direction signals used to guide eye movements. *Journal of Neurophysiology*, *95*(1), 284–300.

Bowns, L. (2011). Taking the energy out of spatio-temporal energy models of human motion processing: The component level feature model. *Vision Research*, *51*, 2425–2430.

Brüderle, D., Petrovici, M., Vogginger, B., Ehrlich, M., Pfeil, T., Millner, S., et al. (2011). A comprehensive workflow for general-purpose neural modeling with highly configurable neuromorphic hardware systems. *Biological Cybernetics*, *104*(4), 263–296.

Burgi, P.-Y., Yuille, A. L., & Grzywacz, N. M. (2000). Probabilistic motion estimation based on temporal coherence. *Neural Computation*, *12*(8), 1839–1867.

Castet, E., Lorenceau, J., Shiffrar, M., & Bonnet, C. (1993). Perceived speed of moving lines depends on orientation, length, speed and luminance. *Vision Research*, *33*(14), 1921–1936.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, *9*(4), 301–315.

Grossberg, S., Mingolla, E., & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. *Vision Research*, *41*(19), 2521–2553.

Hildreth, E., & Koch, C. (1987). The analysis of visual motion: From computational theory to neuronal mechanisms. *Annual Review of Neuroscience*, *10*, 477–533.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95.

Isard, M., & Blake, A. (1998). Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, *29*(1), 5–28.

Lorenceau, J., Shiffrar, M., Wells, N., & Castet, E. (1992). Different motion sensitive units are involved in recovering the direction of moving lines. *Vision Research*, *33*(9), 1207–1217.

Lu, Z.-L., & Sperling, G. (2001). Three-systems theory of human visual motion perception: Review and update. *J. Opt. Soc. Am. A*, *18*(9), 2331–2370.

Majaj, N. J., Carandini, M., & Movshon, J. A. (2007). Motion integration by neurons in macaque MT is local, not global. *Journal of Neuroscience*, *27*(2), 366–370.

Masson, G. S., & Ilg, U. J. (Eds.). (2010). *Dynamics of visual motion processing: Neuronal, behavioral and computational approaches*. Berlin: Springer.

Masson, G. S., Montagnini, A., & Ilg, U. J. (2010). When the brain meets the eye: Tracking object motion dynamics of visual motion processing. In U. J. Ilg & G. S. Masson (Eds.), *Dynamics of visual motion processing* (pp. 161–188). New York: Springer.

Masson, G. S., & Perrinet, L. U. (2012). The behavioral receptive field underlying motion integration for primate tracking eye movements. *Neuroscience and Biobehavioral Reviews*, *36*, 1–25.

Masson, G. S., & Stone, L. S. (2002). From following edges to pursuing objects. *Journal of Neurophysiology*, *88*(5), 2869–2873.

Montagnini, A., Mamassian, P., Perrinet, L. U., Castet, E., & Masson, G. S. (2007). Bayesian modeling of dynamic motion integration. *Journal of Physiology (Paris)*, *101*(1–3), 64–77.

Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1985). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117–151). Rome: Vatican Press.

Nowlan, S. J., & Sejnowski, T. J. (1995). A selection model for motion processing in area MT of primates. *Journal of Neuroscience*, *15*, 1195–1214.

Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science and Engineering*, *9*(3), 10–20.

Pack, C. C., & Born, R. T. (2001). Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain. *Nature*, *409*, 1040–1042.

Pack, C. C., Gartland, A. J., & Born, R. T. (2004). Integration of contour and terminator signals in visual area MT of alert macaque. *Journal of Neuroscience*, *24*(13), 3268–3280.

Pack, C. C., Livingstone, M. S., Duffy, K. R., & Born, R. T. (2003). End-stopping and the aperture problem: Two-dimensional motion signals in macaque V1. *Neuron*, *39*(4), 671–680.

Pei, Y.-C., Hsiao, S. S., & Bensmaia, S. J. (2008). The tactile integration of local motion cues is analogous to its visual counterpart. *Proceedings of the National Academy of Sciences USA*, *105*(23), 8130–8135.

Pei, Y.-C., Hsiao, S. S., Craig, J. C., & Bensmaia, S. J. (2010). Shape invariant coding of motion direction in somatosensory cortex. *PLoS Biology*, *8*(2), e1000305.

Pei, Y.-C., Hsiao, S. S., Craig, J. C., & Bensmaia, S. J. (2011). Neural mechanisms of tactile motion integration in somatosensory cortex. *Neuron*, *69*(3), 536–547.

Perrinet, L. U. (2010). Role of homeostasis in learning sparse representations. *Neural Computation*, *22*(7), 1812–1836.

Perrinet, L. U., & Masson, G. S. (2007). Modeling spatial integration in the ocular following response using a probabilistic framework. *Journal of Physiology (Paris)*, *101*(1–3), 46–55.

Rodman, H. R., & Albright, T. D. (1989). Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Experimental Brain Research*, *75*(1), 53–64.

Rust, N. C., Mante, V., Simoncelli, E. P., & Movshon, J. A. (2006). How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, *9*(11), 1421–1431.

Simoncelli, E. P., & Heeger, D. J. (1998). A model of neuronal responses in visual area MT. *Vision Res.*, *38*(5), 743–761.

Smith, M. A., Majaj, N., & Movshon, J. A. (2010). Dynamics of pattern motion computation. In G. S. Masson & U. J. Ilg (Eds.), *Dynamics of visual motion processing: neuronal, behavioral and computational approaches* (pp. 55–72). Berlin: Springer.

Tlapale, É., Kornprobst, P., Bouecke, J., Neumann, H., & Masson, G. (2010). *Towards a bio-inspired evaluation methodology for motion estimation models* (Tech. Rep. No. RR-7317). Lille: INRIA.

Tlapale, É., Kornprobst, P., Masson, G. S., & Faugeras, O. (2011). A neural field model for motion estimation mathematical image processing. In M. Bergounioux (Ed.), *Mathematical image processing* (pp. 159–179). Berlin: Springer.

Tlapale, É., Masson, G. S., & Kornprobst, P. (2010). Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, *50*, 1676–1692.

Tsui, J.M.G., Hunter, J. N., Born, R. T., & Pack, C. C. (2010). The role of V1 surround suppression in MT motion integration. *Journal of Neurophysiology*, *103*(6), 3123–3138.

Wallace, J. M., Stone, L. S., & Masson, G. S. (2005). Object motion computation for the initiation of smooth pursuit eye movements in humans. *Journal of Neurophysiology*, *93*(4), 2279–2293.

Watamaniuk, S. N., McKee, S. P., & Grzywacz, N. M. (1995). Detecting a trajectory embedded in random-direction motion noise. *Vision Research*, *35*(1), 65–77.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Wilson, H. R., Ferrera, V. P., & Yo, C. (1992). A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, *9*(1), 79–97.

Received September 18, 2011; accepted April 6, 2012.