



**HAL**  
open science

# A unified dynamic neural field model of goal directed eye-movements

Jean-Charles Quinton, L. Goffart

► **To cite this version:**

Jean-Charles Quinton, L. Goffart. A unified dynamic neural field model of goal directed eye-movements. *Connection Science*, 2018, Embodied Neuronal Mechanisms in Adaptive Behaviour, 30 (1), pp.20-52. 10.1080/09540091.2017.1351421 . hal-01637024

**HAL Id: hal-01637024**

**<https://hal.science/hal-01637024v1>**

Submitted on 17 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

To appear in *Connection Science*  
 Vol. 00, No. 00, Month 20XX, 1–32

## A unified dynamic neural field model of goal directed eye-movements

J.C. Quinton<sup>a\*</sup> and L. Goffart<sup>b</sup>

<sup>a</sup>*Laboratoire Jean Kuntzmann, UMR5224, Université Grenoble Alpes / CNRS, France*

<sup>b</sup>*Institut de Neurosciences de la Timone, UMR6193, CNRS / Aix-Marseille Université, France*

(Received 00 Month 20XX; accepted 00 Month 20XX)

Primates heavily rely on their visual system, which exploits signals of graded precision based on the eccentricity of the target in the visual field. The interactions with the environment involve actively selecting and focusing on visual targets or regions of interest, instead of contemplating an omnidirectional visual flow. Eye-movements specifically allow foveating targets and track their motion. Once a target is brought within the central visual field, eye-movements are usually classified into catch-up saccades (jumping from one orientation or fixation to another) and smooth pursuit (continuously tracking a target with low velocity). Building on existing dynamic neural field equations, we introduce a novel model that incorporates internal projections to better estimate the current target location (associated to a peak of activity). Such estimate is then used to trigger an eye movement, leading to qualitatively different behaviors depending on the dynamics of the whole oculomotor system: 1) fixational eye-movements due to small variations in the weights of projections when the target is stationary, 2) interceptive and catch-up saccades when peaks build and relax on the neural field, 3) smooth pursuit when the peak stabilizes near the center of the field, the system reaching a fixed point attractor. Learning is nevertheless required for tracking a rapidly moving target, and the proposed model thus replicates recent results in the monkey, in which repeated exercise permits the maintenance of the target within in the central visual field at its current (here-and-now) location, despite the delays involved in transmitting retinal signals to the oculomotor neurons.

**Keywords:** saccades; smooth pursuit; active perception; dynamic neural field; sensorimotor contingency

### 1. Introduction

Dating back to the ideomotor principle from James (1890) in *Psychology*, the idea that perception and action are tightly intertwined is now widespread and integrated under many different forms in many theories of behavioral and brain sciences. Not only perception guides action, but action also orients perception, both developing and working synergistically for the survival and the maintenance of autonomy of living beings (Barandiaran, 2016). To act adequately in a dynamic environment, as subjectively perceived based on one's own sensorimotor capabilities, either due to motion in the environment or self-generated movement, synchrony must be maintained despite delays in the sensory, motor and neural pathways of information (Buisson & Quinton, 2010). Anticipation may thus be conceived as a key component to correctly interact with the environment and a fundamental principle underlying the brain activity (Bubic, Von Cramon, & Schubotz, 2010).

---

\*Corresponding author. Email: quintonj@univ-grenoble-alpes.fr

Yet, it may actually take on a variety of forms, on a continuum stretching from purely implicit anticipations (relying on past experience to adjust the behavior to the current and present situation) to explicit ones (with representations of predicted future states), as proposed by Pezzulo (2008). While (more or less abstract) goals and intentions exert an influence down to motor primitives (Rosenbaum, Chapman, Weigelt, Weiss, & van der Wel, 2012), it is not trivial to determine how explicit and under which form the anticipatory components are represented. Especially, an internal model predicting the future state of the environment or system may not be required, as interactions with the environment provide the information needed to refine and adjust the course of action, or using Brooks' words (1995): the world is its own best model. The online and implicit differentiation of the various interaction potentialities based on their internal outcomes may be sufficient (Bickhard, 1999). Sensorimotor capabilities, as well as the regularities we learn through interactions, may thus well shape the way the world is perceived. This again goes back to the notion of *Umwelt* from Uexküll (1909) and entails a relative conception of space, where space is constructed by individuals and culture, as already proposed by Poincaré (1914) or developed in Piagetian constructivism (Piaget, 1967).

### 1.1. *Visual perception*

Focusing on visual perception alone and considering primates (including humans), eye movements can thus be generated to actively sample the external environment, and take part in sensorimotor contingencies (O'Regan & Noë, 2001). Part of the complexity of the relationship between an animal and its environment is thus shared between the sensory and motor aspects of visual perception. Indeed, a non-homogeneous visual stimulation where details are only accessible in the central region of the visual field is sufficient to correctly interact with a natural environment, as long as it is complemented by eye movements and coarse peripheral signals to inform about the outcomes of potential gaze shifts. Furthermore, attention, roughly defined as the active selection and filtering of information based on current skills, knowledge, and context, can be deployed at the overt (e.g. eye movements) and covert (e.g. neural modulation) levels.

Adopting a division classically made when modeling attentional systems (Hopfinger, Buonocore, & Mangun, 2000), this covert deployment of attention results from bottom-up or exogenous factors (e.g. stimulus saliency) as well as top-down or endogenous factors (e.g. stimulus relevance for the task at hand). For instance in (Bar, 2007), the magnocellular pathway allows making top-down predictions about possible object categories based on coarse information, which is then complemented by the detailed information provided by the slower parvocellular pathway. Yet, in such a dynamical model of visual perception, even when neglecting eye movements, the frontier between the top-down and bottom-up streams is blurred, as the predictions of course depend on expected and known objects, but also directly depend on the sensory stimulation. Additionally, the projections from the prefrontal cortex (object guesses) are often considered as future-oriented in the literature, but actually simply need to get integrated and thus synchronized with the fine grained details when they reach the relevant areas. Directly adding eye movements to the equation, Kietzmann, Geuter, and König (2011) have experimentally demonstrated that simply manipulating the initial locus of visual attention on ambiguous stimuli was sufficient to causally influence the patterns of overt attentional selection, and the recognition of the object's identity. Once again, the differentiation between alternative percepts seems dynamical, rely on expectations about the stimuli, and the sensorimotor interactions are in fact constitutive of visual

perception. These three ingredients of visual perception (dynamics, sensorimotricity, and anticipation) are the focus of the current paper, modeling both the covert and overt deployment of visual attention.

Since the exact dynamics of eye movements can itself be quite complex, they have been categorized in different types (see for instance Rolfs, 2009). When looking at static scenes, and at the larger spatial and temporal scales, the eye trajectory is mainly composed of saccades (rapid eye movements) separated by fixations. Nevertheless, these are not independent, and are integrated over time for the coherent perception of extended objects to occur, with possibly predictive aspects (Rolfs, 2015). When looking at the spatio-temporal details, the eyes also continuously perform what is called fixational eye movements, which are a combination of drifts, tremors, and micro-saccades. While the role of saccades to explore the visual environment and foveate stimuli of interest has been previously discussed, micro-saccades may similarly contribute to active vision at a different scale (Hicheur, Zozor, Campagne, & Chauvin, 2013). The neural and functional segregation of the different types of eye movements is debated, and it has been argued that common neural mechanisms may underly saccades and micro-saccades generation (Krauzlis, Goffart, & Hafed, 2017; Otero-Millan, Troncoso, Macknik, Serrano-Pedraza, & Martinez-Conde, 2008).

Now returning to dynamic visual environments, eye-movements toward moving targets are also differentiated, with the interceptive saccades (made first toward the target), the catch-up saccades (made when the lag between the gaze and target directions increases), and the slow pursuit eye movements. To unify the different types of eye-movements with a dynamical perspective, the maintenance of a target within the central visual field can be viewed as a dynamic equilibrium within parallel visuo-oculomotor streams (Goffart, Hafed, & Krauzlis, 2012; Hafed, Goffart, & Krauzlis, 2009). When tracking a moving target, the gaze attempts to synchronize its movement with the motion of the target. The idea that an internal model of the target trajectory would guide the gaze direction (e.g., Daye, Blohm, & Lefèvre, 2014) has recently been questioned (Quinet & Goffart, 2015) because of the failure to define the notion of “trajectory” and the lack of explanation how this notion might be represented in the brain activity.

## 1.2. *Dynamic neural fields*

Dynamic neural fields (DNF) are mathematical models based on differential equations, which describe the spatiotemporal evolution of activity (e.g. mean firing rate or membrane potential) in populations of neurons organized within maps. The context of their original development imposed the analytical resolution of the dynamical system, itself requiring strong assumptions on the initial conditions and stimuli characteristics (Amari, 1977; Wilson & Cowan, 1973). Yet, the increasing processing capabilities of computers later allowed simulating the dynamics of such models (Taylor, 1999), at the same time relaxing the constraints and allowing the study of their response to dynamic stimuli.

The core properties of most DNF models is their ability to detect and filter information in the input signal. Targets are detected based on their assumed temporal stability in the signal, as well as their spatial coherence with the lateral connectivity kernel (i.e. the weights of connections between the constituting neurons) (Gepperth, 2014), even in presence of a large amount of noise in the signal. This selection occurs when the system converges to an attractor which encodes the stimulus of interest, under the form of a localized peak of activity (population coding). The parameters and non-linearities

in the equation allow implementing qualitatively different mechanisms, from memory formation to sensorimotor coupling. Also, DNF models do not need to be directly plugged on sensory inputs. Therefore, they are not in principle limited to the detection of stimuli defined by a single lateral connectivity profile (for an instance of complex preprocessing to abstract from raw input images, see Maggiani, Bourrasset, Quinton, Berry, & Sérot, 2016). Also, complex transformations and projections can be considered, for instance using alog-polar representation of the visual field (Taouali, Goffart, Alexandre, & Rougier, 2015) or relying on self-organizing maps (Lefort, Boniface, & Girau, 2011). While Taouali et al. (2015) provide details on how the DNF approach can account for the encoding of target location in the deep superior colliculus, Gandhi and Katnani (2011) review the different plausible decoding mechanisms from a topologically organized population of neurons, focusing on saccade generation. They compare the “vector summation” and “vector averaging” approaches, the latter being preferred in most of the DNF literature, as well as in the current paper. Regardless of their shared characteristics, computational architectures of visual perception and visual attention based on DNFs come in different flavors, based on the way attractors and transitions between attractors are exploited.

On the one hand, the equation can be tuned to switch between different stable attractors when specific conditions are met. For instance, Sandamirskaya (2014) distinguishes detection and selection instabilities (leading to the selection of one or several targets), working memory instabilities (maintenance of a peak even when the target disappears from the input signal), or reverse detection instability (relaxation of the peak of activity). The rationale and biological inspiration is synthesized in Schöner (2008), and this approach has been successfully applied to model behavior and to build a wide range of vision and robotic systems (Schöner & Spencer, 2015). This approach is particularly well suited to model discrete behavioral sequences, where specific conditions must be met before turning to the next step (Sandamirskaya & Schöner, 2010). The same applies for sequences of fixations in static environments, with DNF models of the “what” and “where” components of visual perception (Schneegans, Spencer, Schöner, Hwang, & Hollingworth, 2014), and of the overt and covert deployment of visual attention (Fix, Rougier, & Alexandre, 2010). However, the delays required to build and relax the peaks makes the modeling of tracking eye movements more complicated.

On the other hand, the robustness of the DNF equations in presence of dynamic input can be directly exploited. For slow target motions (projected within the spatial limits of the neural field), the peak drifts and lags behind the target (Rougier & Vitay, 2006). Nevertheless, the system is able to adapt and converge on a limit cycle attractor, tracking the target. The equation can be modified to further improve the tracking accuracy by biasing the peak dynamics in a given direction, either by making the lateral connectivity kernel asymmetric (Cerdeira & Girau, 2010), or by adding internal projections of activity to the stimulation (Quinton & Girau, 2011). The latter is chosen and extended in this paper, as we model the tracking of rapidly moving stimuli, as well as the dynamic transitions between the different types of eye-movements.

In section 2, we start by presenting the original DNF model and its extension for tracking a target moving along a continuous trajectory (PNF). Then, we extend the system and its dynamics, and introduce the novel ANF model, which incorporates eye-movements to account for the fixation and tracking of targets crossing the visual field. We finally add learning capabilities to the system, to extend its adaptability to trajectories of unknown characteristics. In section 3, we again incrementally test the different components of the

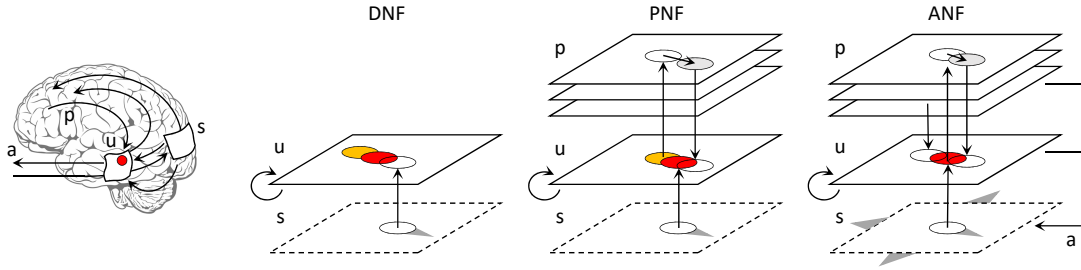


Figure 1. Different neural field models used to detect and track a moving target. The original DNF equation may converge on a limit cycle attractor in presence of the target, yet lagging behind it. On the contrary, the projections of the PNF version allows to converge and synchronize with the target motion. If the target crosses the visual field, only the ANF version is able to smoothly track the target, thus mimicking the dynamics of the brain-body-environment system through visuomotor interactions.

full ANF model, studying the attractors of the associated complex dynamical systems, and evaluating the systems capability to generate different types of eye movements. We then introduce and replicate empirical observations made in monkeys learning to track a target moving along rectilinear trajectories, with dynamic transitions and combinations of saccades and slow pursuit eye movements. We finally close the paper with a short discussion and perspectives.

## 2. Computational model

In order to progressively introduce the components of the ANF model, this section starts from the dynamic neural fields initially proposed by Wilson and Cowan (1973), developed by Amari (1977), and later applied to visual attention directed to moving targets (Rougier & Vitay, 2006). We then introduce the PNF model with a single projection that can bias the dynamics of the neural fields (Quinton & Girau, 2011), then with many projections (Quinton & Girau, 2012), before including eye movements. The increase in model complexity goes hand in hand with extended attentional capabilities, from the detection of fixed targets in the visual field to the generation of tracking eye movements toward a moving target. This progression is illustrated in Fig.1.

### 2.1. Dynamic Neural Field (DNF)

Inspired by neurophysiological studies, including early studies in the visual cortex (Hubel & Wiesel, 1962), dynamic neural fields are mean-field models of the neural activity when it is observed at the mesoscopic scale. The membrane potential reflects the averaged activity of a large number of topologically organized neurons in response to an input stimulus and its local excitability. Formally, the neural field is represented as a 2D manifold ( $\mathbf{X}$ ) in bijection with  $[-0.5, 0.5]^2$ , and described by a distribution of potential  $u$ . The potential at location  $\mathbf{x}$  and time  $t$  is written as  $u(\mathbf{x}, t)$ , and the stimulation as  $s(\mathbf{x}, t)$ . Adopting the single-layer field equation of lateral inhibition type (Amari, 1977), the dynamics of the potential is described by:

$$\tau \frac{\partial u(\mathbf{x}, t)}{\partial t} = -u(\mathbf{x}, t) + c(\mathbf{x}, t) + s(\mathbf{x}, t) + h \tag{1}$$

where  $\tau$  is the time constant,  $h$  the resting potential and  $c(\mathbf{x}, t)$  the component resulting from lateral interactions over the neural field, defined by:

$$c(\mathbf{x}, t) = \int_{\mathbf{x}' \in \mathbf{X}} w(\mathbf{x}' - \mathbf{x}) f(u(\mathbf{x}', t)) d\mathbf{x}' \quad (2)$$

where  $f$  is the activation function (in this paper we opted for a rectified linear unit function), while  $w(\mathbf{x}, \mathbf{x}')$  is the lateral connection weight function satisfying Eq. 3.

$$w(\Delta\mathbf{x}) = A_+ e^{-\frac{\|\Delta\mathbf{x}\|^2}{\sigma_+^2}} - A_- e^{-\frac{\|\Delta\mathbf{x}\|^2}{\sigma_-^2}} \quad (3)$$

In a few words, the equation is composed by a relaxation term (driving the local potential to return to its resting level in the absence of stimulation), a lateral interaction term (leading to the selection of one target), and the external stimulation (input driving the neural field dynamics). With adequate stimuli, the resulting selection of a target is proven to occur when the difference of Gaussians in Eq. 3 is further constrained to a Mexican hat profile, so that the amplitudes and standard deviations satisfy  $A_+ > A_-$  and  $\sigma_+ < \sigma_-$  (Taylor, 1999). In such a case, localized peaks of activity grow on the neural field where stimuli best match the kernel constraints (most often roughly, as for instance with circular visual targets when Gaussian based kernels are used). Additional constraints can be imposed on the parameters such that one single peak can grow and maintain at a given time, especially through the use of global inhibition with a large  $\sigma_-$  value (see Table 1 for the values adopted throughout this paper). The peak location on the neural field (itself supposedly an estimate of a target location) can be estimated as the center of mass  $\mathbf{x}_u$  of the field activity  $\mathbf{u}$  (see Eq.4).

$$\mathbf{x}_u = \frac{\int_{\mathbf{x} \in \mathbf{X}} u(\mathbf{x}) \mathbf{x} d\mathbf{x}}{\int_{\mathbf{x} \in \mathbf{X}} u(\mathbf{x}) d\mathbf{x}} \quad (4)$$

Eq.1 can be simplified by relying on a map based notation for the different components, thus assimilating  $u(\mathbf{x}, t)$  to a  $\mathbf{X} \rightarrow [0, 1]$  map written as  $\mathbf{u}$ , thus leading to the following equation:

$$\tau \frac{\partial \mathbf{u}(t)}{\partial t} = -\mathbf{u}(t) + \mathbf{c}(t) + \mathbf{s}(t) + h \quad (5)$$

Turning to the computational implementation, we rely on an explicit Euler schema to simulate the dynamics. Time is thus discretized with an interval  $dt$  between 2 estimations, and a square regular lattice discretization is used for space (allowing a matrix-based implementation). The continuous version in Eq. 5 becomes:

$$\mathbf{u}(t) = \left(1 - \frac{dt}{\tau}\right) \mathbf{u}(t - dt) + \frac{dt}{\tau} (\mathbf{w} * f(\mathbf{u}(t - dt)) + \mathbf{s}(t)) \quad (6)$$

with  $\mathbf{u}$ ,  $\mathbf{s}$  and  $\mathbf{w}$  being the matrix equivalents for variables and functions  $u$ ,  $s$  and  $w$ . To the exception of a few lines of work (e.g. Cerda & Girau, 2010), the kernel function  $w$  is usually isotropic (as in Eq. 3). The absence of  $\mathbf{x}$  in this formulation reflects the fact that a peak of maximal amplitude only grows in presence of a stationary localized

stimulus, and aligns with it. In response to a stimulus which is not aligned with the peak of activity, the properties of the DNF equation and lateral connectivity kernel lead to an averaging of the stimulus projection, and thus its location, at time  $t$  (with weight  $\frac{dt}{\tau}$ ) with the reentrant feedback of the neural field, with a peak that emerged up to  $t - dt$  (with weight  $1 - \frac{dt}{\tau}$ ). When combining different sources of activity to drive the field dynamics, this averaging property has been widely exploited in the literature, for instance to model saccadic motor planning (Kopecz & Schöner, 1995) in the context of active vision, or multimodal integration models able to replicate the McGurk effect (Lefort et al., 2011). Yet in the context of tracking a moving target, this property implies that the peak necessarily lags behind the input stimulus with the standard equation. The amount of lag depends on the  $dt$  and  $\tau$  parameters, as well as on the stimulus dynamics (e.g. the stimulus instantaneous velocity, that is noted  $\mathbf{v}_s$  in the rest of the paper). Nevertheless, this lag cannot be fully canceled with an isotropic kernel (as illustrated in the *DNF* part of Fig. 1). Additionally, increasing the  $\frac{dt}{\tau}$  ratio enhances the reactivity of the neural field, yet simultaneously reduces its relaxation time and filtering efficiency, two key components of computational models of attention.

## 2.2. Predictive Neural Field (PNF)

To synchronize the peak of activity with the input stimulus, as in presence of a rapidly moving target, the reactivity of the field can be enhanced at the cost of robustness to sensory noise (by adjusting the equation parameters). Yet, to reduce the tracking lag without relying on the trade-off between reactivity and stability, the original equation can be modified in two ways: by using asymmetric kernels, or by turning the stimulation term into a combination of both the sensory signal and an internal model (corresponding roughly to a memory trace) of the stimulus dynamics. Asymmetric kernels have been used in Cerda and Girau (2010) with different motions requiring different kernels ( $\mathbf{w}_k$ ). To discriminate between complex motions, the kernels become location dependent, and iterations over many time steps are required. A set of independent neural fields have been previously used to cover the range of expected motions, but kernels could also be dynamically weighted and combined to reflect the probability of the expected motions (model) to fit the actual dynamics (observation). Here, in order to facilitate transitions between trajectories, and to permit their combination into novel trajectories with which the system can synchronize, we rather proceeded on the decomposition of the stimulation term (Quinton & Girau, 2011, 2012). From a neural perspective, these approaches may all break down to adjusting synaptic weights on the long term (to adapt to new forms of motion), and synaptic gating on the short term (to exploit the repertoire of expected motion). Yet, the modularity and the precise dynamics of these approaches differ because of the way the information is distributed and injected in the equations. In this paper, the term  $\mathbf{s}(t)$  in Eq. 5 is thus replaced by  $\hat{\mathbf{s}}(t)$ , defined as:

$$\hat{\mathbf{s}}(t) = \alpha \mathbf{p}(t) + (1 - \alpha) \mathbf{s}(t) \quad (7)$$

where  $\alpha$  weights the external stimulation  $\mathbf{s}(t)$  and an internal projection  $\mathbf{p}(t)$ . This projection corresponds to the activity required to cancel the lag of the neural field for a specific hypothetically observed trajectory, by exciting the field in the direction of the motion, ahead of the current peak location, and inhibiting the field behind the peak. While it may seem counter-intuitive to make such a substitution, incoming signals for both components may arrive from distal neural areas, only with some of them more



directly connected to the actual physical stimulus (in terms of temporality or similarity). Their origin cannot be accessed in the current model, and due to their symmetrical roles in the equation, they cannot be differentiated at the level of the neural field. Alternatively, adopting a division classically made when modeling attentional systems (Hopfinger et al., 2000), these components may respectively be described as bottom-up and top-down, while the voluntary nature of such influences is beyond the scope of the current paper. The component  $\mathbf{p}(t)$  is further decomposed in Eq. 8, but in this paper, it is exclusively computed from the field potential  $u(t)$ . A null  $\alpha$  restores the original dynamics, while an  $\alpha$  of 1 leads to a closed-loop dynamics disconnected from the external stimulation. An  $\alpha$  in  $]0, .5]$  guarantees that the dynamics is not driven by the internal projections, but by the actual external dynamics.

$$\mathbf{p}(t) = \frac{\sum_k w_k(t) \mathbf{p}_k(t)}{\sum_k w_k(t)} \quad (8)$$

The term  $\mathbf{p}$  is expressed as a combination of individual projections  $\{\mathbf{p}_k\}$ , which correspond to distinct dynamics observed in different contexts (e.g. different trajectories adopted by different targets). The set  $\{\mathbf{p}_k\}$  might as well define a basis of projections, learned from the statistics of observed trajectories, and which can be combined, endowing the system with the ability to adapt to more complex dynamics. Any complex and extended trajectory can for instance be approximated by a smooth interpolation over a sequence of projections, each projection corresponding to a simpler local target motion (temporal weighting). Each unique local projection can also result from a combination of a limited set of projections (spatial weighting), corresponding to target motions with specific directions and speeds. These population coding properties within the neural field have been described elsewhere (for independent projections see Quinton and Girau (2011); for combinations within a set of projections see the supplemental material from Quinton and Girau (2012)). The weight  $w_k$  of each projection changes with time; it is not to be confused with the fixed connection weights  $w$  in the original DNF equation. While the former reflects the adequacy of projection  $\mathbf{p}_k$  with the currently observed dynamics and thus translates the competition between projections in mathematical terms, the latter guarantees the competition between possible locations for the peak growth (as introduced in section 2.1). Although the details on the updating of the weight  $w_k$  are not required at this point, it can simply be considered as a similarity measure between the actual neural field activity  $\mathbf{u}$  and the expected activity (approximated by  $\mathbf{p}$ ). Basically, the associated dissimilarity may be a norm based on point to point differences at the map level, continuously updated during the simulation.

Again, the normalization in Eq.8 guarantees the influence of the overall internal projection of activity on the neural field will not have primacy over the input, since the number of projections on which it is based may vary (e.g. due to learning effects of repeated exercise and the variability of observed trajectories, therefore not reducible to a single projection). Simultaneously high weights for two projections should reflect their similarity, with the neural field averaging their activity. On the contrary, two equally active yet clearly distinct projections will not both lead to a large similarity with the neural field activity (where at most a single peak is present), and a bifurcation will necessarily occur in their weights. Although other aggregation functions could be considered in replacement of the simple linear combination of the projections (Eq. 8), the neural field dynamics attenuates or nullifies the influence of projections with relatively small weights through lateral interactions. In these three instances, we indirectly exploit the usual selection (for distant stimuli), averaging (for nearby target

stimuli and their overlapping field activations) and robustness (for distributed noisy stimulations) properties of the dynamic neural fields.

As previously introduced, projections may be defined for any type of target trajectory, but in this paper, we will limit their scope to rectilinear motions with a constant speed. Stimuli adopting such trajectories will be used in the evaluation section to test the robustness of the model and replicate experimental data. Also, and as any trajectory can be approximated with an arbitrary precision by a set of local linear motions, the aforementioned restriction should be of no consequence for the expressive power of the model. Each projection can be defined as a geometric transformation of the field potential  $\mathbf{u}(t)$ , a transformation that may itself depend upon time, but will be limited here to a fixed translation (which may be combined to approximate more complex trajectories). We thus define:

$$\mathbf{p}_k(\mathbf{x}, t) = u(\mathbf{x} - \Delta\mathbf{x}, t - \Delta t) \tag{9}$$

which corresponds to a shift of  $\Delta\mathbf{x}$  performed in  $\Delta t$ . From a neural perspective, we do not want to consider delay differential equations, in which the continuous propagation of activity over the neural field would lead to infinitesimal shifts when considering decreasing periods of time. If we consider the internal projections to be the result of back-and-forth projections between neural fields or neural areas, there will be a minimal delay  $\Delta t \gg 0$  before any signal generated from the neural field activity returns to it. Then, for any fixed “projection velocity”  $\mathbf{v}_k$ , the equivalent shift must be set, or adjusted, so that  $\Delta\mathbf{x} = \mathbf{v}_k\Delta t$ . The delay  $\Delta t$  may of course depend on the connectivity, thus also on the location within the neural field, but will be considered constant in the current paper with any loss in expressiveness. Also, and to return the mathematics of the system, a large enough delay is required to discriminate between projections and allow bifurcations in the system (through weights  $w_k$ ), since differences will become negligible when the delay approaches zero (in terms of signal-to-noise ratio). Reversely, a too large delay leads to projections no more interacting with a pre-existing peak of activity (through the connectivity kernel of limited excitatory width), and thus has a reduced impact on the neural field dynamics. Practically, there is no need for such projection to be defined for any  $\mathbf{x} \in \mathbf{X}$ , as they simply need to sufficiently cover the neural field, which embeds continuity and smooth out irregularities (with increased weights  $w_k$  possibly compensating for the reduction of effect due to sparsity). Instead of considering projections as transformations of the entire neural field activity, we can indeed decompose projections as local stimulations, responding to distant and past activity on the field. For each location targeted by a projection, we can define an input receptive field on the neural field, reacting to the presence of a peak of activity. Such local projections then define a set of tuning curves (Zhang & Sejnowski, 1999) and the system dynamics will remain roughly the same, as long as the receptive fields overlap sufficiently.

When turning to the computational implementation of Eq. 9, we want to respect both the spatial and temporal discretization of the neural field. We therefore synchronize the updates with  $\Delta t = n \times dt$  ( $n \in \mathbb{N}^*$ ). Additionally, as long as both  $dt$  and  $n$  are neither too small (with bifurcation issues) nor too large (with lateral interaction issues),  $n$  remains a free parameter of the system. It is here arbitrarily set to 1, which gives the following equation when also introducing a second term to facilitate the relaxation of activity at

the previous (peak) location:

$$\mathbf{p}_k(\mathbf{x}, t) = u(\mathbf{x} - \mathbf{v}_k dt, t - dt) - u(\mathbf{x}, t - dt) \quad (10)$$

With this equation, a peak centered on  $\mathbf{x}$  at time  $t - dt$  should move to the location  $\mathbf{x} + \mathbf{v}_s dt$  at time  $t$ . At this stage, the reader might wonder about the relationship between the peak velocity  $\mathbf{v}_u$  (derivative of  $\mathbf{x}_u$ ), the projection velocity  $\mathbf{v}_k$  and the instantaneous velocity of the tracked stimulus  $\mathbf{v}_s$  (of course unknown and not explicitly computed within the neural field model). If all velocities are expressed in the same coordinate system, we should expect the peak to move at speed  $\mathbf{v}_u = \mathbf{v}_s$ , ideally in perfect synchrony with the stimulus (i.e. with no lag). The naive approach is to consider that a single projection  $\mathbf{p}_{k_0}$  from Eq. 8 should be active (i.e.  $w_k \neq 0 \iff k = k_0$ ), or that projections should combine into a single equivalent projection  $\mathbf{p}_k$  associated with  $\mathbf{v}_k = \mathbf{v}_s$ . However, the dynamics of the peak (including its location and shape) makes the picture a bit more complicated. For instance, the projection only contributes to the updated activity with coefficient  $\alpha \frac{dt}{\tau}$ , so that we need to choose  $\mathbf{v}_k = \gamma \mathbf{v}_s$  with  $\gamma > 1$  in order to compensate for the peak inertia (which is a direct consequence of the original equation). A reformulation and approximation of the problem would thus lead to the following definition to optimally synchronize with the target motion:

$$\mathbf{p}_k(\mathbf{x}, t) = u(\mathbf{x} - \gamma \mathbf{v}_s dt, t - dt) - u(\mathbf{x}, t - dt) \quad (11)$$

Further details and demonstration on this speed overestimation are provided in Quinton and Girau (2011), and are illustrated in the *PNF (fast)* part of Fig. 2. In practice, a simple model as expressed in Eq.11 cannot account for the complex dynamics of the neural field, and learning is required to correctly estimate the map between  $\mathbf{u}$  and  $\mathbf{p}$  for any fixed target velocity  $\mathbf{v}_s$ . It is unrealistic to assume targets should follow a trajectory with stable characteristics on a sufficiently large time window  $[0, T]$ , so that a projection  $\mathbf{p}_k$  could be independently learned to synchronize the peak and target dynamics. In practice, the set of weights  $w_k$  should be also learned as functions of the neural field state, testing if the projection facilitates the synchronization with the target, for which the actual location, velocity, or even presence are unknown. While a form of predictive coding could be considered, the crucial problem remains that the projection velocity  $\mathbf{v}_k$  and the peak velocity used to test the prediction are different, and related by a complicated function that also depends on most of the neural field parameters. When combining all these factors, the problem is far from being trivial to solve through unsupervised learning. These difficulties will not be directly addressed in this paper, and are simply highlighted as they will partly be dissolved when turning to the ANF model in the next section. We will therefore rely on Eq. 10 in the rest of the paper.

With target motions remaining within the limits of the visual field, the aim of the PNF model is to extend the covert attentional capabilities of classical DNFs. As described in Quinton and Girau (2011), the field potential  $\mathbf{u}$  may converge to a limit cycle in presence of a periodic target trajectory, even when the stimulus speed is increased, and without the alternation of decaying and building peaks observed with the original equation. That is, the peak velocity  $\mathbf{v}_u$  on the neural field will be equal to the stimulus velocity  $\mathbf{v}_s$ . Also, robustness to noise and distractors is increased, since the target is now characterized by its spatiotemporal dynamics and not only by its shape compatibility with the kernel profile. Thus, and contrary to Rougier and Vitay (2006) for instance, distractors can have the exact same spatial characteristics and be visible during the entire simulation,

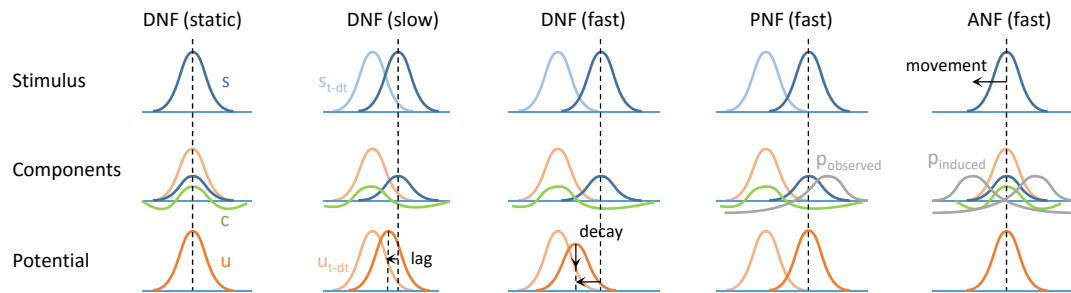


Figure 2. Illustration of the neural field model dynamics and equation components. For each model, stimulation at time  $t - dt$  (light blue) and  $t$  (blue), field activity from  $t - dt$  (light orange) and resulting at  $t$  (orange), as well as the competition component (green), are represented. Internainputl projections corresponding to expected trajectories are also shown for the PNF and ANF models. With a moving stimulus, the DNF model is not able to converge to a stable peak of maximal activity, the peak lagging behind the target and/or decaying. The projection of activity in the PNF model simply compensates for the inertia of the equation and synchronizes with the target trajectory (if the implicit expectation and observation match). The ANF model additionally includes a projection due to the eye movement, thus canceling the drifting effect of the projection when the target image is kept within the central visual field.

but simply adopt a non-expected trajectory. It is therefore possible to discriminate and interpolate between trajectories in a nonlinear fashion, thanks to the feedback loop of projections validity ( $w_k$ ) in the PNF equation (Quinton & Girau, 2012). Finally and as a consequence, the system becomes also robust to the temporary disappearance of the target (e.g. due to occlusion). In such a case, the peak decays in absence of stimulation, but initially keeps tracking the stimulus at its current (here and now) location. Indeed, the inertia induced by the original equation remains effective, with the peak and projection positively exciting each other (a high  $w_k$  leading to a larger peak amplitude and vice versa). Nevertheless, this behavior is short-lived (a few hundred milliseconds with current parameters) since the input and thus the weight of other projections rapidly regain a significant impact (because of  $\alpha < .5$  in Eq.7).

In the context of this paper, the most striking feature of the PNF model is that the projection of activity  $\mathbf{p}_k$  on the neural field is actually not a prediction of the present state  $\mathbf{u}(t)$ . Prediction usually supposes a projection into the future, and while the projection  $\mathbf{p}_k$  is indeed built based on the state of the field activity at time  $t - \Delta t$  (and thus oriented toward the future), it only influences the dynamics and effectively gets projected onto the field at time  $t$  (thus in the present, if we consider the external world to be setting the time reference). More importantly, while the field potential  $\mathbf{u}(t - \Delta t)$  can be conceived as the input to generate an internal expected or experienced trajectory, the projection  $\mathbf{p}_k$  generated in return is a form of internal action on the neural field, which alters its dynamics. The projection term thus is not required to be in direct relation with the external stimulation  $\mathbf{s}$ , but simply allows maximizing the synchrony between the system and stimulus dynamics, i.e. minimizing the distance between  $\mathbf{x}_s$  (actual stimulus location) and  $\mathbf{x}_u$  (estimated location). In other words, we should not expect the weight  $w_k$  (somehow the confidence in the projection ability to facilitate synchrony) to be correctly estimated based on the difference between the projection  $\mathbf{p}_k$  and the present field state  $\mathbf{u}(t)$  or stimulation  $\mathbf{s}(t)$ . Even though they end up being combined in the neural field space  $\mathbf{X}$ , they correspond to qualitatively different concepts; if  $\mathbf{u}(t - \Delta t)$  provides part of the context,  $\mathbf{p}_k$  is the action, and  $\mathbf{u}(t)$  the outcome. At this point, two limitations lead to the development of the ANF model: 1) Such projections require to learn or approximate the relation between  $\mathbf{v}_s$ ,  $\mathbf{v}_u$  and  $\mathbf{v}_k$  for any location on the field. 2) The model is limited

to covert attention (with the absence of eye movements), as a target leaving the visual field will be lost.

While a heterogeneous field of view (e.g. characterizing the retina with a central macula and fovea) could be viewed as a complicating factor for the above issues, it is actually a solution to the problem when one considers the generation of eye movements. Foveating a target indeed facilitates its pursuit by increasing the neural resources for processing the motion signals, and at the same time hinder and make unnecessary the learning of stimuli dynamics for all locations of the visual field. Learning to foveate the target and keep it in the central visual field will reduce the problem to processing the motion signals and characterize the dynamics of the input stimulus. Also pointing in such direction, a model of target position encoding in the superior colliculus has been recently proposed in Taouali et al. (2015), demonstrating that apparently complex foveation performance can be accounted for by dynamical properties of neural field and retinotopic stimulation (using a log-polar representation of the visual field).

### 2.3. *Active Neural Field (ANF)*

The new Active Neural Field model proposed in this paper extends the previous models by introducing eye movements, complementing the covert attentional deployment (Quinton & Girau, 2011), with overt shifts of attention. The system thus aims at reaching a fixed point attractor when a target is stabilized in the visual field, while performing eye movements synchronized with the motion of the target. Action is now not only performed through projections to the neural field, but also through a physical movement of the simulated eyes. To simplify the model, we will here rely on a Cartesian coordinate system and a homogeneous resolution over the visual field, although some advantages of turning to a more realistic retinal model will be developed in the final discussion. While such increments to the model may seem limited in terms of equations, they drastically extend the capabilities of the system, change the characterization of the attractors of the associated dynamical system, as well as the dimensionality and structure of the learning space. The following sections therefore focus on these aspects.

Physical actions are initiated once an activity threshold  $u_a$  is reached on the neural field, reflecting the detection of a target. With a sufficiently high value for  $u_a$ , the maximal activity of the neural field ( $\|u\|_{max}$ ) should indeed correspond to the center of an existing or forming peak, and the field activity can be directly exploited to orient the simulated eye toward the estimated target location. An alternative way to achieve the same result, often adopted in the DNF literature, is to adjust the resting level  $h$  from the original DNF equation, while setting  $u_a = 0$ . A peak maintained above the threshold will continuously feed the movement generators, but such a peak can hardly be maintained away from the neural field center (details will be provided with the analysis of the system dynamics in the result section). This threshold reflects a kind of “cost” for moving the eyes (energy, time, blur), since its absence would lead to continuous movements of possibly large amplitude. Indeed, in absence of an already high peak, the estimated (peak) location (following Eq. 4) can be subject to quick variations, simply because of small variations of the potential  $u$  induced by noise in the input signal. The activity threshold  $u_a$  is set to .4 in this paper, but is dependent on the neural field parameters.

All actions in this system are aimed at bringing the target within the central visual field (generating an associated peak at the center of the neural field), consequently

limiting the parameter space to explore for learning. Actions can thus be represented as a map  $\mathbf{a}$  from the field activity to the motor system. In turn, the action  $\mathbf{a}$  leads to an apparent movement on the field of view  $-\mathbf{x}_a$ , which is approximately equal to  $-\mathbf{x}_u$  (thus the movement simply zeroes out the eccentricity of the target at its current estimated location). In experimental data, both smooth pursuit and saccadic eye movements correspond to continuous movements, but with qualitatively different speed profiles. Although testing the fit between the generated saccade profiles and those observed in experimental settings is meaningless here, due to the simplistic implementation of oculomotor control, the apparent discontinuity of the movement in the model is merely due to the temporal discretization of the differential equations. As a proof of concept model and for simplicity and explanatory purpose, saccades will always be performed over one time step in the current model, regardless of their amplitude. As a consequence, and for such movements to remain biologically plausible, a lower limit to  $dt$  is imposed for the simulation (this is already guaranteed when considering the projection delay  $\Delta t$ , as described in previous section).

To refine the previous notations for a controllable field of view, let  $\mathbf{x}_f$  be the center of the field of view in the world reference frame. Assuming Cartesian coordinates, and applying the adequate action  $\mathbf{a}(\mathbf{u}(t))$  to center the currently focused target should lead to the following changes at  $t + dt$ :

$$\mathbf{x}_f(t + dt) = \mathbf{x}_f(t) + \mathbf{x}_a(\mathbf{a}(\mathbf{u}(t))) = \mathbf{x}_f(t) + \mathbf{x}_u(t) + \eta \quad (12a)$$

$$\mathbf{x}_s(t + dt) = \mathbf{x}_s(t) - \mathbf{x}_a(\mathbf{a}(\mathbf{u}(t))) = \mathbf{x}_s(t) - \mathbf{x}_u(t) - \eta \quad (12b)$$

where  $\eta$  is an error term that is minimized through learning to allow perfectly centered fixations of targets. While the action map  $\mathbf{a}$  and the sensory consequence  $\mathbf{x}_a$  of performing the action can both be easily given here (starting from Eq.4 for the extraction of the peak location), such transformations can be learned in DNF models. For instance, Bell, Storck, and Sandamirskaya (2014) focus on saccade generation and learning of the coupling between target location encoding and motor commands. Even with the simple Cartesian coordinate systems used here, the world and visual field reference frames must be clearly distinguished, or equivalently the apparent location  $\mathbf{x}_s$  of the stimulus from its physical location in the world reference frame, defined by  $\hat{\mathbf{x}}_s = \mathbf{x}_f + \mathbf{x}_s$ . With the emergence of a peak in response to a static target, we get  $\mathbf{x}_u(t) \approx \mathbf{x}_s(t)$  before the saccade is initiated, and thus  $\mathbf{x}_s(t + dt) \approx 0$ . The learning of eye movements can be made independent to that of the projections dealing with observed movements, as the structure of the visual space can be acquired when looking at or away from static targets (or those adopting slow movements for which the original DNF dynamics is sufficient). Because of the inertia in the original equation, projections can also be associated to eye movements, facilitating the relaxation of the eccentric peak and the emergence of a new (theoretically central) peak of activity. The associated projection of activity can thus be defined by:

$$p_a(\mathbf{x}, t) = u(\mathbf{x} - \mathbf{x}_a(\mathbf{a}(\mathbf{u}(t - dt))), t - dt) - u(\mathbf{x}, t - dt) \quad (13)$$

The entire neural field architecture for the ANF model is represented on Fig. 3. When a peak of activity grows on the neural field in response to a target, saccades, once performed, systematically induce stimulations near the center of the visual field. For the

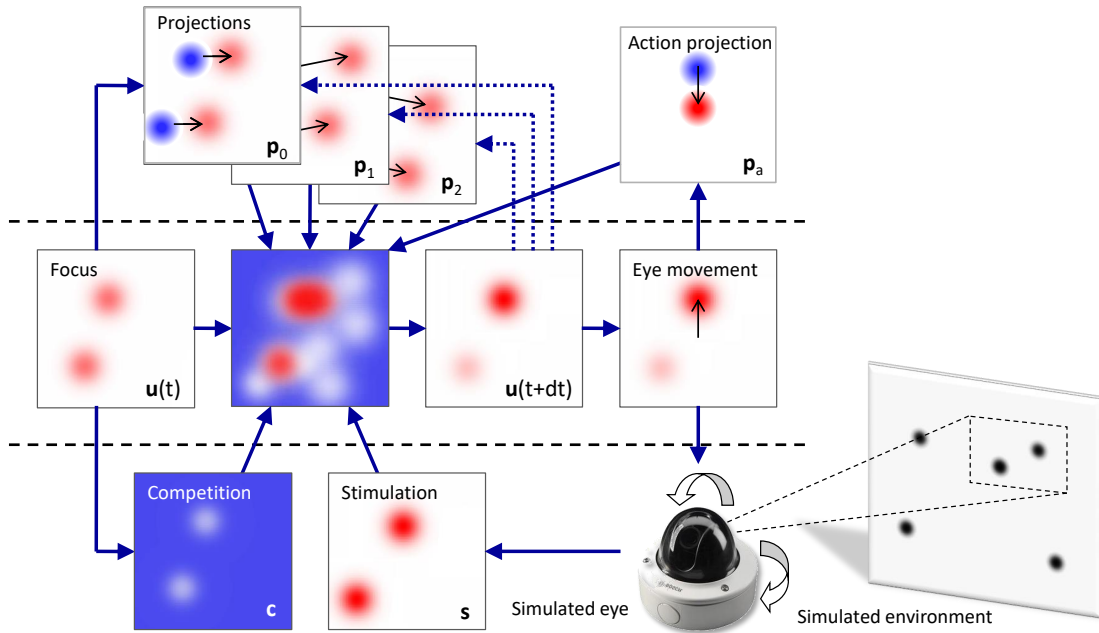


Figure 3. Architecture of the Active Neural Field (ANF) model. The model first includes the lateral interaction ( $c$ ) and integration of input stimulation ( $s$ ) from the DNF model. It also integrates the projections ( $p_k$  with weight  $w_k$ ) from the PNF model, themselves evaluated based on their ability to predict the current (here-and-now) location of the target ( $\hat{p}_k$ ). In addition, the model also controls eye movements ( $a$ ), and thus can predict the associated transformations on the visual field to be facilitated by a projection of activity ( $p_a$ ) on the neural field. At each simulation time step, all components are combined with the neural field potential  $u(t)$  to produce  $u(t + dt)$ .

previously introduced PNFs, the prediction  $\hat{p}_k$  and projection  $p_k$  in Eq. 10 and 11 will not be correct or at least optimal anymore, as they are designed for target motions with constant velocities. Yet, if a moving target leads to a peak away from the center, the projection associated with the expected motion ( $p_k$ ) will not be compatible with the post-saccadic central stimulation, while both the inhibitory component from the  $p_k$  and  $p_a$  components will facilitate the relaxation of the earlier peak and thus a refocus on new apparent location of the target. If the peak appears near the center of the visual field, with the target moving away from the center (i.e. with the eye movement lagging behind the target), projections  $p_a$  and  $p_k$  will have opposite effects on the dynamics. For a projection in adequation with the stimulus velocity,  $p_a$  and  $p_k$  will cancel out and stabilize the peak when the continuous eye movements (induced motion  $x_a$  at each iteration) match the target motion (as illustrated on the ANF part of Fig. 2).

This attractor results from the combined effect of both  $p_a$  and  $p_k$  projections, as well as the stimulation, when they are symmetrical near the position  $v_s dt$  in the neural field reference frame. The peak may thus only be stabilized in the direction of the target motion and slightly away from the neural field center. Despite the resulting small lag, this allows at the same time to continuously test the validity of different projections, while limiting the need to learn regularities in the entire visual field. An accurate alignment between the gaze and target directions would be possible by directly associating the eye movements with projections, and by simultaneously estimating the similarity between the performed and proposed movement, as well as drifts in the peak location (reflecting a mismatch between the expected and actual

target motion). Such an approach based on sensorimotor contingencies in the context of active vision has been used in Quinton, Catenacci, Barca, and Pezzulo (2014) and Catenacci, Quinton, and Pezzulo (2014), but the subsequent increase of dimensionality of the learning space is better suited to well-structured stimuli. Here on the contrary, a small lag behind the target (at least momentarily) makes the decoupling of the target foveation and the oculomotor behavior possible, since the validity of the associated projections can be estimated when the peak grows or shifts slightly away from the center of the neural field. As a consequence,  $\mathbf{p}_k$  and  $\mathbf{p}_a$  can be acquired separately, and as the peak is expected to converge to a fixed location in the visual field, the problem of estimating the effect of  $\mathbf{p}_k$  on the peak movement in the PNF model dissolves.

To conclude this section, adopting an active vision approach by coupling eye movements with the neural field dynamics yields two main positive consequences: 1) learning the regularities becomes simpler, as only the central part of the visual field becomes critical for the accurate tracking of a target (a process which would in fact be facilitated by a non-homogeneous visual stimulation), 2) inertia in the equation still allows to filter out noise and distractors, but the constraint on target speed is reduced, as physical motor actions now replace peak movements (the system converges to a sensorimotor fixed point attractor, instead of a limit cycle in the visual space alone).

#### 2.4. *Simulation parameters and performance metrics*

Before turning to the results, and to facilitate their reproducibility, we here specify the model parameters and exact simulation scenarios. As the ANF model is equivalent to the PNF model in the absence of eye movements and associated projections, and the PNF model to a DNF model by only considering projections with  $\mathbf{v}_k = \mathbf{0}$  (instead of setting  $\alpha = 0$  as previously described), we used the exact same parameter values for all simulations. The description, symbol and value (or range of values) for all parameters from the DNF, PNF and ANF equations are reported in Table 1. Although the number of parameters may seem elevated, they are not independent for a given objective behavior (such as tracking a target). For instance, increasing the time constant  $\tau$  facilitates noise filtering, but for the system to remain sufficiently reactive to track a rapidly moving target, the kernel usually needs to be simultaneously adjusted. For any given parameters, increasing the speed of the moving target, the number of distractors or the level of noise in the signals sooner or later breaks the expected performance of the system. Parameters have here initially been optimized using genetic algorithms for maximizing the tracking performance of the DNF model (Quinton, 2010). This way, and by not retuning the parameters for the PNF and ANF models, we opt for the worst case conditions, where performance improvement due to the model extensions are minimized and can be hardly be explained by confounding factors.

As the objectives of this paper deal with detecting and tracking a moving target through different types of eye movements, we do not include distractors in the stimulation, and thus simply present a single noisy stimulus defined by:

$$\mathbf{s}(\mathbf{x}, t) = \hat{f} \left( A_s e^{-\frac{\|\mathbf{x}_s - \mathbf{q}\|^2}{\sigma_s^2}} + \epsilon \right) \quad (14)$$

with  $\epsilon$  being a white noise component sampled in  $\mathcal{N}(0, \sigma_\epsilon)$ , while  $\hat{f}$  is a saturation



Table 1. Parameters and variables of the neural field model.

Group of parameters	Description	Symbol	Value/Space
Visual field	Visual field space	$\mathbf{X}$	$[-0.5, 0.5]^2$
	Coordinates within the field	$\mathbf{x}$	$\in \mathbf{X}$
	Field coordinates (world reference frame)	$\mathbf{x}_f$	$\in \mathbb{R}^2$
Neural field equation / potential		$\mathbf{u}$	$\mathbf{X} \rightarrow [0, 1]$
	Time constant	$\tau$	0.2
	Resting potential	$h$	0
	Firing / activation function (ReLU)	$f$	$\mathbb{R} \rightarrow \mathbb{R}^+$
	Stimulation / projection coefficient	$\alpha$	0.5
Lateral connectivity kernel	Focus coordinates (peak center)	$\mathbf{x}_u$	$\in \mathbf{X}$
		$\mathbf{w}$	$\mathbf{X} \rightarrow [-1, 1]$
	Excitatory amplitude	$A_+$	0.06
	Excitatory range (sd)	$\sigma_+$	0.1
	Inhibitory amplitude	$A_-$	0.03
Gaussian stimulus	Excitatory range (sd)	$\sigma_-$	1.0
		$\mathbf{s}$	$\mathbf{X} \rightarrow [0, 1]$
	Stimulation amplitude	$A_s$	$\in \{0, 1\}$
	Stimulation range (sd)	$\sigma_s$	0.1
	Stimulus noise (white)	$\epsilon$	$\mathbf{X} \rightarrow \mathbb{R}$
	Noise amplitude (sd)	$\sigma_\epsilon$	0.2
	Stimulus velocity	$\mathbf{v}_s$	$\in [-3, 3]^2$
Projection(s)	Stimulus location (field reference frame)	$\mathbf{x}_s$	$\in \mathbf{X}$
	Stimulus location (world reference frame)	$\hat{\mathbf{x}}_s$	$\in \mathbb{R}^2$
		$\mathbf{p}_k$	$\mathbf{X} \rightarrow [-1, 1]$
	Projection weight (validity)	$w_k$	$\in [0, 1]$
Eye movement	Projection “velocity”	$\mathbf{v}_k$	$\in [-5, 5]^2$
	Motor space	$\mathbf{A}$	$[-1, 1]^2$
	Motor command	$\mathbf{a}$	$\mathbf{A}$
Discretization / dynamics	Induced spatial displacement	$\mathbf{x}_a$	$[-1, 1]^2$
	Associated projection	$\mathbf{p}_a$	$\mathbf{X} \rightarrow [-1, 1]$
	Activation threshold	$u_a$	0.4
Discretization / dynamics	Time	$t$	$\in \mathbb{R}$
	Time step	$dt$	0.05
	Spatial resolution	$n$	51

function of unity slope (rectified linear unit), simply guaranteeing the stimulation values remain within  $[0, 1]$ .  $\mathbf{x}_s$  is the center of the Gaussian target profile, as used for instance in 12b, whose trajectory varies between results subsections. The dispersion of the stimulus distribution  $\sigma_s$  could in theory be a free parameter, but should nevertheless be constrained by initial visual preprocessing steps for the tracking system to perform efficiently. As stated in Geppert (2014), DNFs “contain a *data model* which is encoded in their lateral connections, and which describes typical properties of afferent inputs. This allows inferring the most likely interpretation of inputs, robustly expressed through the position of the attractor state”. Therefore,  $\sigma_s$  is here taken equal to  $\sigma_+$  to converge on near Bayesian optimal decisions.

To characterize the performance of the computational system, we rely on two complementary measures. First, an error signal estimates the distance  $d_e$  between the center of the field of view  $\mathbf{x}_f$  and the actual target location  $\mathbf{x}_s$ . Since one cannot expect the system to focus on the target before it has been presented, the time variable is adjusted to guarantee that the target should be visible for any  $t \in [0, T]$ , with  $T$  corresponding either to the disappearance of the target, or the end of the simulation. Thus, we can

estimate the mean error distance  $\overline{d_e}$  as:

$$\overline{d_e} = \frac{1}{T} \int_0^T \|\mathbf{x}_f - \mathbf{x}_s\| dt \quad (15)$$

Assuming that the target location varies during the entire simulation (either continuously when tracking a single moving target, or with discontinuities when switching between static targets), this error increases drastically when the target is lost (since  $\mathbf{x}_f$  will be no more matching  $\mathbf{x}_s$ ). Also, it is inflated when the peak lags behind the target (positive  $\|\mathbf{x}_f - \mathbf{x}_s\|$  distance) or if the peak relaxes before focusing at a new position (with a maximal distance just before a new saccade is performed).

The characterization of saccadic movements usually relies on a shift in signal frequency. In the case of experimental data on eye movements, they usually correspond to stereotyped profiles with speeds reaching up to  $1000^\circ/s$  (in humans and monkeys) and lasting only a few milliseconds ( $< 200\text{ms}$ ). Most eye-tracking hardware or analysis software rely on delay or speed thresholds to classify movements into saccades and fixations. Because of the temporal discretization of the system used in the simulations, the movement speed is at best estimated over a single time step ( $dt$ ). The saccade, catch-up saccade and smooth pursuit movement distinction thus qualitatively depends on the instantaneous speed of the movement. Although small movements can be generated at each time step as long as the activity over the field remains high enough (over the action threshold), large amplitude movements (of high speed, and thus saccadic) cannot be sustained over several time steps. The same dynamical system thus generates: 1) continuous movements with an adaptive instantaneous speed defined as the derivative between successive locations, 2) saccadic movements, which induce a large change in field stimulation, and thus require the field to relax over several time steps (through lateral inhibition) before another movement can be generated (once the activity reaches again the action threshold).

To more accurately describe the performance of the system, we want to differentiate intervals during which smooth pursuit movements are made with a small lag, from series of small amplitude saccades. Following the previous paragraph, and from the dynamical system perspective adopted in this paper, the fundamental difference comes from the non-linearity introduced by the action threshold, with smooth tracking simply corresponding to a maximal field potential maintained above  $u_a$  over several time steps. Relying on the temporal discretization of the equations, we thus define a saccade-like movement to occur whenever  $a_{\text{saccade}} = 1$ , following Eq.16.

$$a_{\text{saccade}}(t) = \begin{cases} 1, & \text{if } \|u(t)\|_{\text{max}} \geq u_a \text{ and } \|u(t - dt)\|_{\text{max}} < u_a \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

For the purpose of differentiating smooth pursuit and saccadic eye movements when they hold the same error statistics  $\overline{d_e}$ , we introduce a second measure as the number of saccades performed during the entire simulation. Numerically, this is computed as the sum of  $a_{\text{saccade}}(t)$  for  $t \in [0, T]$ . The time period is set to be consistent with Eq.15, but no saccade should be performed outside this time window when no target is presented. Combining the two measures is therefore necessary, as losing a target only increases  $\overline{d_e}$ , since no saccade is performed afterwards, thus lowering the relative number of saccades performed as time passes when compared to an accurate tracking behavior. The prototypical combinations of the tracking statistics and their interpretation are synthesized in

Table 2. Tracking statistics and their interpretation based on the type of target dynamics

Target dynamics	Tracking error $\overline{d_e}$	Nb of saccades $\sum_t a_{\text{saccade}}(t)$	Behavior interpretation
(all)	High	High	Random saccades (depending on noise/parameters)
Saltatory movement	Low	Low	Stable fixation on target
—	Low	High	Micro-saccades around target
—	High	Low	Continuous drift away from target
Smooth movement	Low	Low	A few catch-up saccades followed by smooth pursuit
—	Low	High	Catch-up saccades during the entire simulation
—	Medium	Medium	Alternation of fixations and saccades
—	High	Low	Target lost during the simulation

Table 2.

### 3. Results

The ANF model will be tested on different scenarios, to demonstrate its performance, analyze its fine-grained dynamics, and facilitate an incremental understanding of its behavior. The following section thus cover targets with fixed positions over an extended period of time (thus reflecting how the visual system focuses on different features by alternating saccades and fixations), and targets moving along a continuous trajectory (thus testing the generation of catch-up saccades and smooth pursuit).

#### 3.1. Fixational eye-movements

To evaluate the performance of the ANF model, we start by exploring the fine-grained characteristics of the simulated oculomotor dynamics on series of fixations. Using Eq. 14 to generate the sensory stimulation,  $\mathbf{x}_s$  is now sampled in the uniform distribution  $\mathcal{U}(-.5 + \sigma_s, .5 - \sigma_s)$ , thus guaranteeing that a well-defined peak of activity grows within the limits of the neural field, in response to a target present in the visual field. This avoids the need for a log-polar representation of the visual field (favoring the emergence of central peaks) or the use of a toric manifold (which is often adopted in computational studies to escape the boundary problems from the differential equation, e.g. discussed in Rougier & Vitay, 2006). The location of the input stimulus remains the same for periods of 1 second, before jumping to a new random position (sampled from the same distribution). Although an adaptation mechanism responsible for perceptual fading, an inhibition of return mechanism, or top-down modulations of activity should be implemented to reflect realistic dynamics of saccades and fixations, as well as to replicate the usual frequency of saccades (e.g. during free viewing or cognitive tasks), this is not the focus in the current paper and has been explored elsewhere (Quinton et al., 2014). The current system dynamics nevertheless parallels computational models of visual attention based on saliency maps (Itti & Koch, 2001), where the target selection is based on lateral inhibition in recurrent neural networks in the original models (Itti, Koch, Niebur, et al., 1998).

The resulting sequence of fixations generated by the ANF model can be observed in Fig. 4(a), which is a projection of the visual field center along the horizontal axis (i.e. first component of the  $\mathbf{x}_f$  vector) as a function of time. The complete trajectory of the target may cover a large part of the environment, captured by a succession of saccades

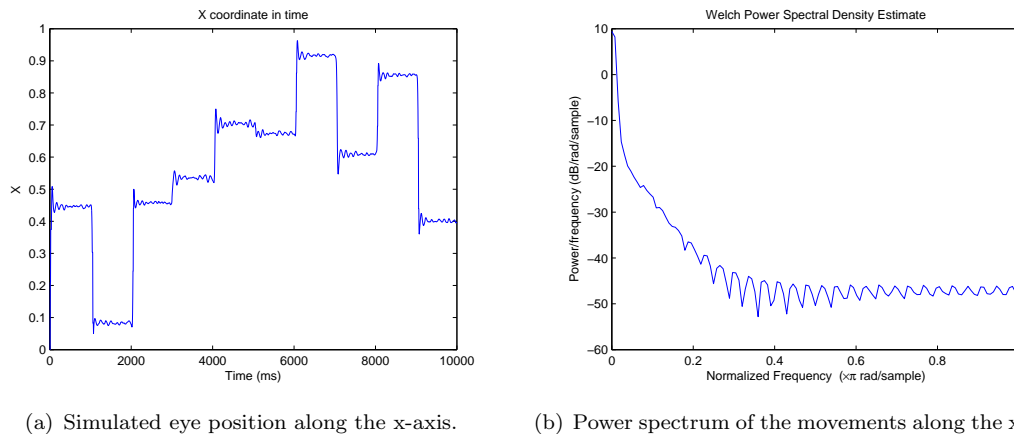


Figure 4. Representative trajectory of the Cartesian visual field coordinates in the environment, when focusing on targets adopting a new random position every second. The left figure reflects the relative complexity of the trajectory, with a sequence of saccades and fixations matching the stimulus dynamics, but also random fluctuations and micro-saccades. The right figure displays the power spectrum of the movements along the x-axis, with a decreasing

whose amplitude yet never exceeds half the size of the visual field (due to the constraints introduced in the previous paragraph). The uniform distribution used to sample the target location leads to a random-walk behavior, with 10 fixations visible in Fig. 4(a).

The oscillations visible in Fig. 4(b) are in part due to the temporal discretization of the differential equation combined with the choice of bandwidth resolution for the Welch power spectral density estimate, but also reflect the closed-loop interactions between the projections and neural field for a fixed target location. Nevertheless, and despite the stochasticity of the simulation, a decreasing power as frequency increases is systematically found. As stated by Astefanoaei, Creanga, Pretegianni, Optican, and Rufa (2014), “using a log-linear scale the visualization of the power spectrum of either random or chaotic data may be a broad spectrum but only chaotic signals are expected to emphasize a coherent decrease of the square amplitude with the frequency increase”. Following the methodology they used when studying how human execute saccadic and fixational eye movements in response to red spots visual stimuli, we qualitatively replicate human behavior (Fig. 2 and 4 of their paper respectively corresponding to the trajectory and power spectrum on the horizontal axis, i.e. Fig. 4(a) and Fig. 4(b) here). Studying the movements separately for each fixation, the power spectral density demonstrates the same decreasing trend. Also, adopting the same statistics to characterize the temporal dependency in the eye trajectory, we found high values for the Hurst exponent ( $M = .70$ ,  $SD = .03$ ). Combined, these results demonstrate that the micro-movements performed by the system during fixations are neither purely oscillatory (where peaks should be observed on the power spectrum) nor purely fully driven by the stimulation noise (since we should find  $H \approx .5$  for Brownian motion).

Although such descriptive statistics based on frequency distributions do not account for all characteristics of fixational eye movements, they are compatible with the frequency shift approach generally used to characterize the different types of eye movements (as introduced in previous section). Without a more refined model of eye movements, we cannot expect the current system to correctly approximate more than the frequency distribution of fixational eye movements (e.g. their spatial distribution). Indeed, motion is only induced by eye movements in this evaluation scenario, and projections are thus

defined by Eq.13. In absence of other projections that may distort space locally, and in presence of white noise in the external stimulation, the projected activity is also statistically anisotropic (the robustness of neural fields to such perturbations being one of their strength). Variability in the stimulation thus influences the estimation of target location, but do not introduce the asymmetries that are found in real visual environments and that would be amplified by the activation of context dependent projections (see next sections). As a consequence, and for static targets, the same pattern is observed for each fixation, and would be for any arbitrary projection direction (x-axis on Fig. 4). With this model in mind, we can argue that fixations are the result of a balanced activity between opposing commands and associated projections. The equilibrium might not be perfect at all time, thus leading to fluctuations around the target location, but the dynamics of the system implies that the target will define a fixed point attractor. Thus, visual fixation can be conceived as a dynamic equilibrium in the visuo-oculomotor system, as proposed by neurophysiological studies (Goffart et al., 2012; Krauzlis et al., 2017).

In addition to the frequency distribution, we can also look at the movements generated by the system once the initial oscillations driven by the non-linear dynamics have settled, i.e. when the fixation conceived as an ideally fixed attractor has been reached. Such movements mainly appear a few hundred milliseconds after a peak has grown in the new central position, since the strong lateral inhibition of the current model prevents several peaks to be maintained near the action threshold ( $u_a$ ). This naturally replicates the delay observed in humans from fixation offset to micro-saccades production, rather independently of the task performed (Otero-Millan et al., 2008). Even though micro-movements performed by the system during fixations (i.e. when the theoretical target location is stable) are partly driven by the sensory noise, the system responds through a coherent behavior, with fluctuations in projection weights influencing the neural dynamics and generated eye movements. Although these movements are only the by-product of a target centering oculomotor behavior yielded by fluctuating signals, they also contribute to locate the target within the visual field, increasing potentially the robustness of the gaze-related signals which guide the movements of e.g., the hand. More careful testing would be required to conclude on this part, but they may also contribute to increase the acuity of the system (see Ko, Poletti, and Rucci (2010) for related human experiments; Franceschini, Chagneux, Kirschfeldand, and Mcke (1991) for early experiments in the fly; Viollet (2014) for a review of visual sensors hyperacuity based on micro-eye movements). They may also contribute to cognitive demands at end (Hicheur et al., 2013) through the activation of task-dependent projections. Generalizing to other movements and equilibriums, a connection can be made with the slight and chaotic movements performed by the human body for keeping balance (Milton, Insperger, & Stepan, 2015).

### 3.2. *Attractor dynamics on a moving target*

In this section, we turn to stimuli crossing the visual field with a constant velocity. For the system to detect and track the moving target, we generated targets which followed rectilinear trajectories defined by:

$$\hat{\mathbf{x}}_s = \mathbf{x}_f + \mathbf{x}_s = \mathbf{v}_s t \quad (17)$$

where  $\hat{\mathbf{x}}_s$  is the target location in the world coordinate system, which differs from the target position within the visual field coordinate system as it depends on  $\mathbf{x}_f$ . Starting the simulation at a negative time  $t \ll -0.5/\mathbf{v}_s$  (but for  $\mathbf{v}_s = \mathbf{0}$ ) and setting the initial

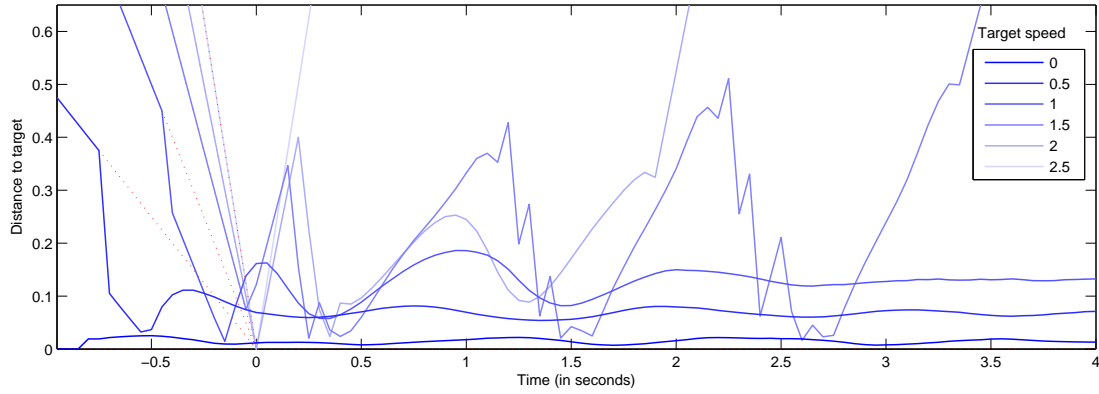


Figure 5. Distance between the visual field center and actual target coordinates ( $\mathbf{d}_e$ ) as a function of time, for a system with a null projected velocity ( $\mathbf{v}_k = \mathbf{0}$ ) and targets moving along a rectilinear trajectory at various speeds ( $\|\mathbf{v}_s\| \in \{0, .5, 1, 1.5, 2, 2.5\}$ ). This allows testing the behavior of the original DNF model endowed with the ability to generate eye movements. All stimuli are designed to reach the center of the visual field at  $t = 0$  in absence of eye movements. *Dotted lines* correspond to the extrapolated target trajectories in the field reference frame, where no eye movement would be performed. For small increasing speeds, the target is detected and tracked with an increasing lag. For larger amplitudes, the target is captured and tracked through sequences of catch-up saccades. For extreme increasing speeds, the target gets lost sooner, or not even detected (for  $\|\mathbf{v}_k\| = 2.5$ ). Values increasing beyond .5 betray a loss of target (i.e. target exiting the visual field that cannot be captured again).

visual field position at  $\mathbf{x}_f = 0$ , the position of the target at  $t = 0$  should be  $\mathbf{x}_s = 0$  (i.e. the center of the visual field). This equation thus guarantees the trajectory crosses over the visual field and passes through its center, as long as no eye movements away from the initial position are produced by the system before the target enters the visual field. Although more complex trajectories could be used, they would be of no consequence for the explanatory power of the model. Also, this section additionally serves the purpose of introducing the results from the following section, in which we replicate empirical observations made in monkeys tracking a target adopting similar trajectories.

In this section, the system includes an additional projection corresponding to a constant velocity ( $\mathbf{v}_k$  in previous sections), thus related to Eq. 10. As previously, we also consider a correct map has been learned for the eye movements (associated with  $\mathbf{p}_a$  from Eq. 13) to center any target located anywhere in the visual field. Fig. 5 illustrates a few representative dynamics by plotting the distance between the visual field center and the actual target location (metrics  $\mathbf{d}_e$ ) as a function of simulation time, running 5 seconds simulations. They represent particular instances in the target speed continuum ( $\|\mathbf{v}_s\| \in [0, +\infty)$ ) with a projection corresponding to a static target ( $\mathbf{v}_k = \mathbf{0}$ ). This configuration allows evaluating the behavior of the original DNF model endowed with the generation of eye movements. With the specific delay parameter chosen, we indeed obtain  $\mathbf{p}_k = \mathbf{u}(t - dt)$ , so that the dynamics of the original equation is maintained, yet slightly altered depending on the value of  $\alpha$  in Eq. 7. The dynamics on the figure can be classified as follows:

- For a static stimulus ( $\mathbf{v}_s = \mathbf{0}$ ), a peak grows at the center of the field after 200ms (since  $\mathbf{x}_s(t) = \mathbf{0}, \forall t$ ), followed by small and slow movements in part driven by stimulation noise (leading to a very small yet non-zero distance).
- For stimulus speed values in  $[.5, 1]$ , the target is detected before it reaches the field center, generating an interceptive saccade to it (drop in error distance around  $t = -750$ ms for  $\|\mathbf{v}_s\| = .5$ , and  $-400$ ms for  $\|\mathbf{v}_s\| = 1$ ).

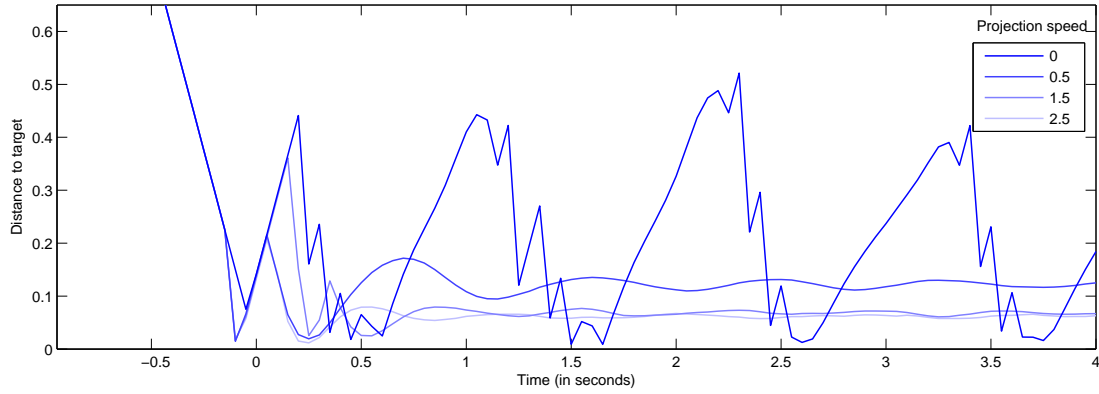


Figure 6. Distance between the visual field center and actual target coordinates ( $d_e$ ) as a function of time, for a system presented with targets moving along a rectilinear trajectory with a constant speed ( $\|\mathbf{v}_s\| = 2$ ), but with different projected velocities, all matching the motion direction of the target, but with different amplitudes ( $\|\mathbf{v}_k\| \in \{0, .5, 1.5, 2.5\}$ ). All stimuli are designed to reach the center of the visual field at  $t = 0$  in absence of eye movements. As the projected speed gets closer to its optimal value, fewer catch-up saccades are performed, and the target is tracked with a diminishing lag. As discussed in the main text and in order to correctly estimate the validity of projections, the lag is never fully canceled.

- For stimulus speed values in  $[1.5, 2]$ , the target passes through the field and is captured before exiting at  $t \approx 200\text{ms}$ , the initial saccade being followed by an alternation of drifts and catch-up saccades, before the target is lost at  $t \approx 3300\text{ms}$  for  $\mathbf{v}_s = 1.5$  and  $t \approx 1900\text{ms}$  for  $\mathbf{v}_s = 2$ .
- For extreme speed values ( $\mathbf{v}_s \geq 2.5$ ), the target just crosses the field in less than 500ms, so that no peak emerges and the target has not been detected in any of the simulations.

Fig. 6 illustrates complementary dynamics, when stimulus speed is fixed at  $\|\mathbf{v}_s\| = 2$  and the projection speed increases, again for a system with a single projection for expected (observed) movements. For a projection corresponding to a target expected to be static ( $\mathbf{v}_k = \mathbf{0}$ ), the behavior observed on Fig. 5 for  $\|\mathbf{v}_s\| = 2$  is roughly reproduced. Yet, the fact that the target is now tracked during the entire simulation (instead of being lost at  $t \approx 3300\text{ms}$ ) reflects the chaotic nature of the system. The precise time of interception and the number of catch-up saccades also vary. Thus, the boundaries between the categories drawn previously must be considered as fuzzy and dependent upon the exact stimulation dynamics.

For projection speed values in  $].5, 2.5]$ , the target is intercepted earlier, centered through a smaller number of catch-up saccades, and then tracked with a decreasing lag when the projection speed gets closer to the actual target speed. The direct comparison of the projection and target speed is made possible by the simple coordinate transformations involved here, yet optimal tracking never occurs for perfectly matching speeds. Small oscillations still appear, reflecting that the relaxation of a slightly lagging peak facilitates its movement, then leading to its maximal activity building up again (see Quinton & Girau, 2011, for additional details in the context of the PNF model). The system thus performs the expected behavior of classical DNFs when presented with static targets, by converging to a stable attractor in the visual space, yet performing eye movements to correctly compensate the target elevated speed.

To get a broader view of the dynamics, and to be able to obtain statistically meaningful results, repeated simulations have been performed to average out the stochastic

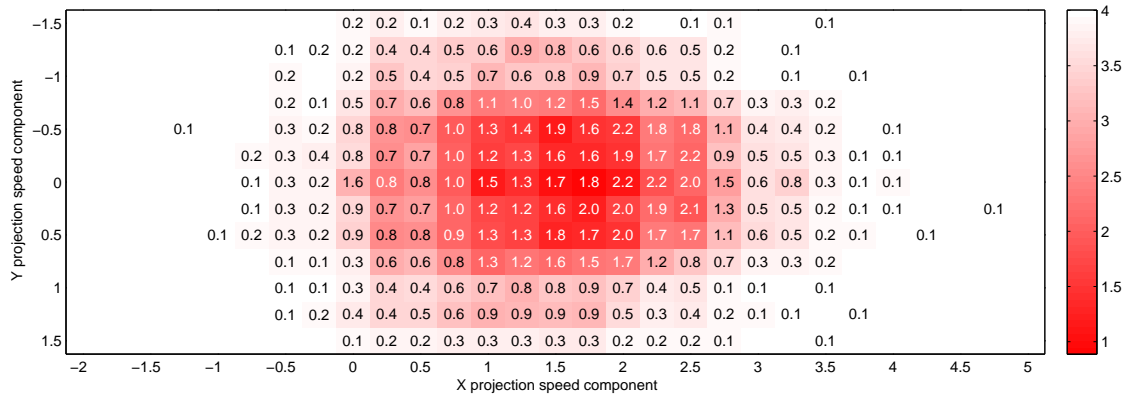


Figure 7. Mean distance between the simulated eye and real target coordinates after  $t = 0$  (*color gradient*) and average number of saccades per trial (*figures*), for a target moving at  $\|\mathbf{v}_s\| = 2$  units/second along the horizontal axis (1 unit = visual field width). Each cell corresponds to 200 simulations of the system running with a single projection, with the corresponding projection velocity components indicated along the axes. A bell-shaped profile appears around the target speed coordinates (with low error, and less than 2 catch-up saccades to capture the target), reflecting the discrimination capabilities of the projections.

components of the equations. We can then characterize the conditions under which the system is able to detect and track the target, using both the error distance ( $\mathbf{d}_e$ ) and number of saccades ( $\sum a_{\text{saccade}}$ ), varying both  $\mathbf{v}_s$  and  $\mathbf{v}_k$ . 200 runs of 5 seconds have been executed for each combination of  $\mathbf{v}_s$  and  $\mathbf{v}_k$  (fixed throughout each simulation). Fig. 7 represents the performance of the system when the stimulus adopts a horizontal velocity set at  $\mathbf{v}_s = (2, 0)$ , and the projection speed  $\mathbf{v}_k$  varies in  $[-2, 2] \times [-1.5, 1.5]$ . The gradient (both in color for error distance and figures for the number of saccades) reflects the optimality of performance for projection speed close to the target actual speed and the very progressive loss of performance when parameters deviate. Yet, in accordance with earlier PNF results and due to the neural dynamics, the optimal configuration does not correspond to  $\mathbf{v}_k = (2, 0)$  but to  $\mathbf{v}_k \approx (1.5, 0)$ , with less than two (interceptive and catch-up) saccades to definitely capture (foveate) the target. Indeed, the dispersion in both metrics is very low for near optimal configurations. The number of saccades here never reaches high values due to the excessive speed of the target, rapidly lost without the proper projection (please refer to Table 2 as a guide for the combined color and figure interpretation). The variability and decimals in the average number of saccades performed reflect the stochastic nature of the process. Values of 0.1 may for instance correspond to targets which have been intercepted in 5% of the simulations (10/200), but where immediately lost.

To test the performance of the system on slow targets (i.e. checking that the performance does not degrade under weaker constraints), and to test the limits of the system with targets adopting extreme speed values, the synthesis of complementary simulations is shown on Fig. 8. While both target and projection velocity this time adopt the same direction, their amplitude both vary, with  $\|\mathbf{v}_s\| \in [0, 3]$  and  $\|\mathbf{v}_k\| \in [-2, 5]$ . The behavior can again be classified as follows:

- $\|\mathbf{v}_s\| = 0$  (first row with static stimulus): Even with extreme projection speeds, convergence on target occurs within one or two saccades, and the tracking error remains low. This demonstrates that the internal projection does not overcome the input stimulus.
- $\|\mathbf{v}_s\| \in (0, 1.25]$  (low target speed): The projection hinders the correct tracking of



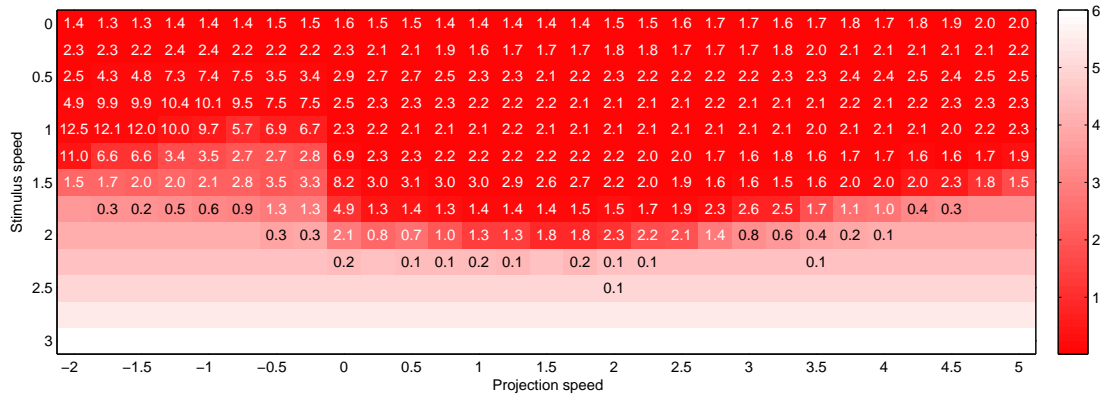


Figure 8. Mean distance between the simulated eye and real target coordinates after  $t = 0$  (*color gradient*) and average number of saccades per trial (*figures*), for a target moving at different speeds and for various projection speeds (both velocity vectors remaining collinear). Each cell corresponds to 200 simulations of the system running with a single projection, with the target and projection speed values indicated along the axes.

the target if the projection velocity is opposed to the target motion. Even though the system is still able to track the target with a relatively low error, the number of catch-up saccades increases drastically when the target speed increases (up to 2.5 saccades/second), highlighting the need for correct projections when they become necessary for capturing the target.

- $\|\mathbf{v}_s\| \in (1.25, 2]$  (high target speed): The target is most often lost with incorrect projections, while performance improves with smaller lag and fewer catch-up saccades for target and projection speeds which are compatible. Symmetrically, the target is lost for excessive projection speeds. The results for  $\|\mathbf{v}_s\| = 2$  statistically match those from Fig. 7 for horizontal projection velocities (row with vertical component equal to 0). The non-linearity of the system can be observed on the figure, as there the relation between the target speed and equivalent optimal projection on the neural field is clearly nonlinear:  $\mathbf{v}_k \approx 3$  for  $\mathbf{v}_s = 1.5$ , thus facilitating the relaxation of a too stable peak, while  $\mathbf{v}_k \approx 1.5$  for  $\mathbf{v}_s = 1.75$ , thus reinforcing a peak that already tends to disappear.
- $\|\mathbf{v}_s\| \in (2, +\infty)$  (extreme stimulus speed): The target is not efficiently tracked in any of the simulations. At best, a few interceptive or catch-up saccades are performed before the target is lost (leading to average numbers of saccades below 1, and very high tracking error).

In all of these simulations, and for the selected tracking statistics, we find that the ANF model outperforms the original DNF model endowed with the generation of eye movements (see configuration  $\mathbf{v}_k = (0, 0)$  on Fig. 7 and  $\|\mathbf{v}_k\| = 0$  on Fig. 8). Integrating all previous results, the dynamics of the system as a function of the stimulation and projections can be simplified as a graph represented on Fig. 9, where nodes correspond to specific dynamics or attractors of the system, and transitions to changes (or maintenance) in projection or target speed. The transitions can be described as follows:

- (A) No spatiotemporally coherent stimulation is provided to the system, and the field potential  $\mathbf{u}$  remains close to its resting level (fluctuations being due to sensory noise).
- (B) A target crosses the visual field, but is yet to be detected and intercepted by the system.
- (C) The target is detected (relatively) early due to its low to medium speed, or thanks to the facilitatory effect of a compatible projection. Nevertheless, the peak always builds with a small lag relatively to the actual target position.

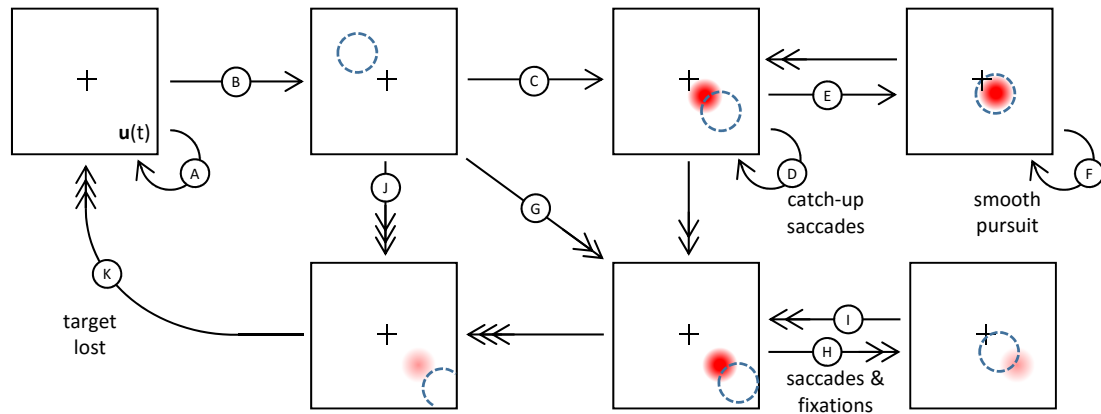


Figure 9. Illustration of the ANF model dynamics in presence of a moving target following a rectilinear trajectory with constant speed. Adopted states and transitions depend on the target dynamics and internal projection of activity, and are thus implicitly represented here using arrows: low or compatible speeds with *single-headed arrows*, high speeds and/or medium discrepancy with *double-headed arrows*, extreme speeds or high discrepancy with *triple-headed arrow*. Apparent target position in the visual field is represented as a *blue dashed circle* while the neural field peak is represented by a *red spot* whose intensity reflects the peak amplitude. Please refer to the main text for the detailed description of the labeled transitions.

- (D) As soon as a peak builds above the threshold for eye movement generation ( $u_a$ ) and away from the center of the visual field, the movement brings the target to a closer location from the center. As the target continues in the same direction, the peak is either continuously reinforced or rapidly reaches the threshold again if it is slightly relaxed, hence generating another catch-up saccade.
- (E) If the peak progressively approaches the position where the projection matches the target here-and-now location, the peak is maintained at a maximal level of activity, and a stable fixed point attractor is reached.
- (F) As the peak activity is maintained above the eye movement threshold, continuous smooth pursuit movements are generated.
- (G) If the target crosses the field at a higher speed (vs. (C) above), the peak will only build later, just before the target leaves the visual field.
- (H) An interceptive saccade of greater amplitude is generated. The peak will then immediately decay below the threshold, since the target now stimulates the center of the field while the decaying peak of course remains in its previous location.
- (I) The relaxation of the peak (or at least its diminishing amplitude) allows a new peak to emerge, when the target is again about to leave the visual field. This alternation generates a sequence of saccades and fixations, which can simply be viewed as an amplified dynamics compared to transition (D).
- (J) If the target crosses the field at an extreme speed (at least relatively to the equation time constant and to (C) and (G) above), no peak is fully formed, and the field activity does not reach the eye movement threshold before the target leaves the visual field. Similar transitions occur from several other states if the stimulus speed increases or if random fluctuations are sufficient to destabilize the peak or delay its formation. With a more general projection model, this should happen when the target adopts an unexpected movement at a sufficiently high speed.
- (K) Once the target is lost and any activity relaxes on the neural field, the system returns to its initial idle state, since no active exploration of the environment is implemented here.

### 3.3. Learning with catch-up saccades and smooth pursuit

In all previously described results, we have assumed the projection was fixed so that trials were totally independent. Here, we aim at replicating observations made by Bourrelly, Quinet, and Goffart (2013) and Bourrelly, Quinet, Cavanagh, and Goffart (2016), demonstrating in macaque monkeys that the ability to accurately pursue a moving target is not spontaneous but must be learned. Over the course of several trials and training sessions, the monkeys improve their ability to foveate the moving target with slow eye movements of progressively increasing velocity while at the same time, the number of catch-up saccades is reduced, as if the foveation system attempted to synchronize the movement of the eyes with the current here-and-now location of the target. We also want to test whether the model fits other empirical results obtained by Quinet and Goffart (2015), showing how the monkey brain differently extrapolates transient target motions depending upon the rate of change in speed.

As a consequence, we try here to qualitatively reproduce *in silico* the paradigm used in these experiments. A fixed target is initially presented at  $\mathbf{x}_s = (0, -0.3)$  from  $t = -1$ s. This artificial fixation point is removed before  $t = 0$  (leaving 200ms for the peak to relax), and replaced by a target moving horizontally with a constant velocity  $\mathbf{v}_s = (1.4, 0)$ , starting from  $\mathbf{x}_s = (0, 0)$ . The speed amplitude of 1.4 is chosen in order to make it possible for the system to track the target in the absence of correct projection, even though this may often require several catch-up saccades, as the system demonstrates a large variability in performance for this configuration (see Fig. 8). At  $t = 0.8$ s, the target disappears, allowing to test the continued dynamics of the system without stimulation. For every trial, the exact same stimulation dynamics is reproduced, except for the sensory noise. During each trial, the eccentricity of the peak is estimated, reflecting the lag of the system and providing a 2D gradient towards the optimal projection (and thus predicted velocity). We thus again assume all mappings between the neural field activity and eye movements and projections have been learned, reducing the dimensionality of the problem without much changing its complexity and characteristics. Nevertheless, convergence relies on the convexity of the 2D problem, which is illustrated on Fig. 7 with the tracking error following a bell-shaped profile based on the 2D projection velocity. The projection (associated to the projection velocity) is arbitrarily updated at the end of each trial, in order to make the connection with previous sections simpler, and to guarantee a correct estimation of the gradient over a sufficient period of time. Since we have previously noted that the peak should grow at an eccentric location reflecting the projection speed, we can directly update the estimate iteratively as follows:

$$\mathbf{v}_k = (1 - \beta) \times \mathbf{v}_k + \beta \times \frac{\overline{\mathbf{x}_u}}{dt} \tag{18}$$

which is a time discretized version of a differential equation with relaxation and input from the current trial estimate.  $\beta$  is set to .05 in the following simulation, which corresponds to a 40s integration time constant for trials of 2s. The estimate of the average peak eccentricity  $\overline{\mathbf{x}_u}$  is computed for  $t \in [0.8, T]$ , in order to let enough time for the peak to emerge, and eliminate the influence of the static target presented at the beginning of each trial. Of course, a more complete version of the model should learn to predict the dynamics of the entire trial (e.g. exact timing at which the moving target appears), but this would no more correspond to the simple case of linear movements studied here.

Fig. 10(a) displays the evolution of  $\mathbf{v}_k$  components over the course of 1000 trials.

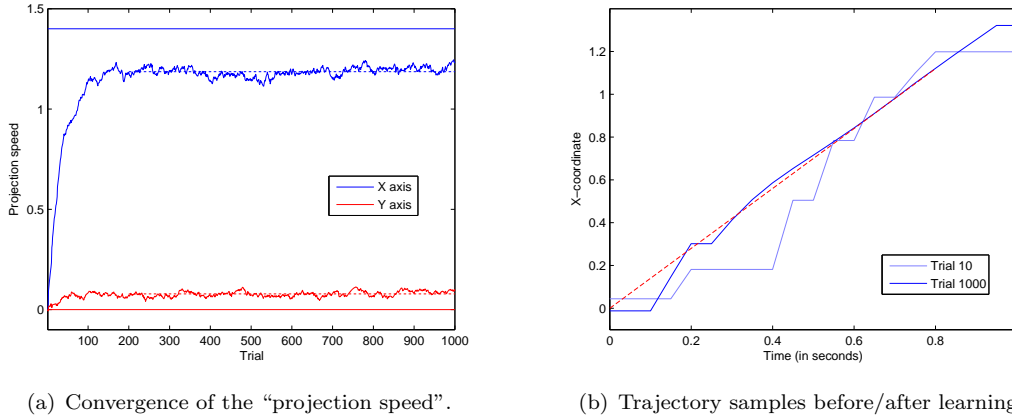


Figure 10. Illustration of the learning dynamics of the ANF system when updating the projection of activity on the neural field after each trials, reproducing experimental data from Bourelly et al. (2016). (a) The projection speed components (*blue* for X, *red* for Y) rapidly converge to relatively stable values (*dashed lines* for the asymptotic means), yet different from the target speed components (*plain lines*), due to the complex neural field dynamics. (b) Representative trajectories for trials #10 and #1000. The behavior drastically changes, with an earlier capture of the target, a reduction in the number of catch-up saccades, the dominance of smooth pursuit, and the continuing tracking at the here-and-now position of the target even after its disappearance (*red dashed line* corresponding to the target trajectory, and stopping at  $t = .8s$ ).

Although convergence occurs rapidly, the continuing variations are due to the signal/noise ratio in the model and the chaotic nature of the performance, limiting the precision of the estimate with the current integration scheme from Eq. 18. Fig. 10(b) illustrates the dynamics of the ANF system on two representative trials, at the beginning of the simulated experiment (trial #10) and at the end (trial #1000). Table 3 also reproduces the two statistics used throughout this paper for the previously mentioned trials, as well as the peak eccentricity (translating the lag of the system in spatial terms). At the end of the learning period, the mean tracking error is highly reduced much below the peak dispersion  $\sigma_+$ , which equals to .1 in this paper. The eccentricity of the peak for trial #1000 is reaching the exact value for which  $\|\mathbf{v}_s\| = \|\dot{\mathbf{x}}_u\|/dt$  ( $1.4 = 0.07/0.05$ ), but does not match the system projection speed for the same trial ( $\mathbf{v}_k = (1.23, 0.09)$ ). Also, the number of saccades drops to 2, with an interceptive saccade followed by a single catch-up saccade. The eye movement is almost perfectly synchronized with the target motion. This remains true for a few milliseconds after the target disappears and until the relaxation of the peak at  $t \approx 950ms$ , demonstrating the robustness and inertia of the PNF equation, here transferred into the sensorimotor system of the ANF model. A decrease in eye movement speed is observed in the empirical data once the target has disappeared, but is not present in the current simulations. This is simply because a single projection is considered in this part of the simulations; without competition between projections corresponding to different hypotheses, and in absence of a spatiotemporally coherent stimulus, the peak dynamics is only driven by the single active projection, at least until the peak relaxes below the action threshold. We thus qualitatively replicate the dynamics and evolution of visual tracking reported by Bourelly et al. (2016, 2013), while the system also statistically lags behind the target, thus also demonstrating that no prediction of the future location of the target is needed for the system to track the target at its current (here and now) location. For individual trials (like #10), there might be a few exceptions, but these are due to noise, incorrect target location estimates, or excessive overestimation of the target speed at the early stages of the learning process.

Table 3. Performance statistics for representative trials during learning

Run	#10	#1000
Mean peak eccentricity ( $\overline{x_u}$ )	0.193	0.070
Mean tracking error ( $\overline{d_e}$ )	0.085	0.027
Number of saccades ( $\sum a_{\text{saccade}}$ )	5	2

#### 4. Conclusion and discussion

Using the formalism of dynamic neural fields theory, this paper brings a proof-of-concept model which is able to replicate different types of eye movements (rapid saccade or slow pursuit) that primates make in response to a visual target, static or moving. More importantly, the qualitatively distinct eye movements emerge as attractors in a sensorimotor dynamical system, without altering the neural field parameters. Once the usual parameters are set, transitions between the two types of eye movements occur depending on the target eccentricity and speed, the amount of noise, as well as the *a priori* knowledge about the possible target trajectories. Slightly altering the neural dynamics through delayed projections of activity is indeed often sufficient to switch from one eye movement type to the other, and to outperform the original DNF model endowed with the same mechanism for eye movement generation. For instance, anticipating the target motion may transform a limit cycle attractor (corresponding to series of saccades with an oscillating peak on the neural field) into a fixed-point attractor in the sensorimotor system (corresponding to a smooth pursuit tracking behavior), by reducing the peak lag and eccentricity of the target in the visual field. Put differently, our results underline the role of both embodiment and timing, since the very same sensorimotor dynamical system leads to apparently distinct visuomotor behaviors, depending on the precise interactions between the dynamics of the neural field, eye movements and target motion.

Although variations of the neural field parameters have not been tested in this paper, most configurations allowing attentional properties to emerge in classically DNF models can be exploited in the PNF and ANF contexts. Also, even though variations in the parameters would alter the results, we also indirectly tested the additional flexibility provided by the addition of projections and eye movements. By learning the optimal projection the system should use to perfectly synchronize with rapidly moving targets, we demonstrated how the system may compensate for an initial lack of performance and for possible inter-individual differences in neural dynamics.

Relatively to the task at hand, parameters are usually adjusted by considering the trade-off between the adaptability and stability of the system (e.g. by preventing or facilitating the relaxation of the peak of activity). This imposes limitations on the level of dynamics the system can cope with (e.g. target maximal speed), thus favoring the modeling of behaviors which can be characterized by a convergence on fixed attractors followed by instabilities, for instance with saccadic eye movements, by relying on temporary inhibition of the neural field activity to switch from one attended location to another. Of course, the intrinsic limitations of neurally plausible field dynamics remain, e.g. by not artificially reducing the time constant in the original equation while increasing the temporal resolution of the simulation. Yet, these limitations are here transcended by the continuous coupling and mutual influence with the motor components. Instead of relying on phase transitions between stable states driven by predictions, observed or induced changes in the sensory flow, we turn to a unique sensorimotor dynamical system in which stable states are not exclusive of movement.

In the near future, the generalization of the results to more complex motions should be tested, relying on the transition and interpolation between expected local trajectories. Target trajectories with a wide range of shapes and dynamics can be computationally studied, for instance aiming at replicating the empirical results on accelerating or decelerating targets (Bourrelly, Quinet, & Goffart, 2014). Further developments are yet required to better model and explain some key principles of the oculomotor dynamics. As briefly discussed in previous sections, the integration of an adaptation mechanism responsible for perceptual fading would allow reflecting realistic dynamics of saccades and fixations. Also, the hypotheses about the functional role of fixational eye-movements (here induced by noise or projections) could be explored by activating or deactivating various components of the model. On a related note, this line of work could extend and bridge the gap with systems which also rely on dynamical systems and dynamical neural fields to model perceptual decision-making and fixation patterns on structured visual stimuli (Quinton et al., 2014), as well as to detect and track objects of interest in presence of combined target-motion and ego-motion (Maggiani et al., 2016).

Finally, a preference for orienting gaze shifts that bring the target near the center of the visual field was here assumed. Explanatory reasons for favoring such saccades can be computationally evaluated, for instance by adopting a retinotopic yet non-homogeneous representation of the visual field (with a greater acuity in the fovea, at the center of the retina), as developed in Taouali et al. (2015). In addition to limiting the extent of the sensorimotor regularities to be learned, a saccade then also leads to the magnification of the target image. This over-representation of the target in terms of neurons can also be reversely seen as a local distortion of space-time, slowing down the apparent motion of the target relatively to the neural dynamics. Such mechanism and those developed in this paper are complementary, both facilitating the tracking of the target for given neural field parameters. These mechanisms should be integrated and their relative contributions evaluated in the future.

## Funding

The research leading to these results has received funding from the French program "investissement d'avenir" managed by the National Research Agency (ANR), from the European Union (Auvergne European Regional Development Funds) and from the "Région Auvergne" in the framework of the IMobS3 LabEx (ANR-10-LABX-16-01), as well as from the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

## References

- Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2), 77–87.
- Astefanoaei, C., Creanga, D., Pretegiani, E., Optican, L., & Rufa, A. (2014). Dynamical complexity analysis of saccadic eye movements in two different psychological conditions. *Romanian reports in physics*, 66(4), 1038.
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, 11(7), 280–289.
- Barandiaran, X. E. (2016). Autonomy and enactivism: Towards a theory of sensorimotor autonomous agency. *Topoi*, 1–22.

- Bell, C., Storck, T., & Sandamirskaya, Y. (2014). Learning to look: A dynamic neural fields architecture for gaze shift generation. In *International conference on artificial neural networks* (pp. 699–706).
- Bickhard, M. H. (1999). Interaction and representation. *Theory & Psychology*, 9(4), 435–458.
- Bourrelly, C., Quinet, J., Cavanagh, P., & Goffart, L. (2016). Learning the trajectory of a moving visual target and evolution of its tracking in the monkey. *Journal of Neurophysiology*.
- Bourrelly, C., Quinet, J., & Goffart, L. (2013). Equilibria and transitions during visual tracking : Learning to track a moving visual target in the monkey. In *Society for neuroscience*. The Association for Research in Vision and Ophthalmology.
- Bourrelly, C., Quinet, J., & Goffart, L. (2014). Unsupervised dynamic morphing of a spatiotemporal visual event during its oculomotor tracking. *Journal of Vision*, 14(10), 492–492.
- Brooks, R. A. (1995). Intelligence without reason. *The artificial life route to artificial intelligence: Building embodied, situated agents*, 25–81.
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in human neuroscience*, 4, 25.
- Buisson, J.-C., & Quinton, J.-C. (2010). Internalized activities. *New Ideas in Psychology*, 28(3), 312–323.
- Catenacci, N., Quinton, J.-C., & Pezzulo, G. (2014). How active perception and attractor dynamics shape perceptual categorization: a computational model. *Neural Networks*(60), 1-16. doi:
- Cerda, M., & Girau, B. (2010). Bio-inspired visual sequences classification. In *Brain Inspired Cognitive Systems (BICS)* (p. pp. 20). Madrid Spain.
- Daye, P. M., Blohm, G., & Lefèvre, P. (2014). Catch-up saccades in head-unrestrained conditions reveal that saccade amplitude is corrected using an internal model of target movement. *Journal of vision*, 14(1), 12–12.
- Fix, J., Rougier, N., & Alexandre, F. (2010). A dynamic neural field approach to the covert and overt deployment of spatial attention. *Cognitive Computation*, 1-15.
- Franceschini, N., Chagneux, R., Kirschfeldand, K., & Mcke, A. (1991). Vergence eye movements in flies. In *Göttingen neurobiology report* (Vol. 1, p. 275).
- Gandhi, N. J., & Katnani, H. A. (2011). Motor functions of the superior colliculus. *Annual review of neuroscience*, 34, 205–231.
- Gepperth, A. (2014). Processing and transmission of confidence in recurrent neural hierarchies. *Neural processing letters*, 40(1), 75–91.
- Goffart, L., Hafed, Z. M., & Krauzlis, R. J. (2012). Visual fixation as equilibrium: evidence from superior colliculus inactivation. *The Journal of Neuroscience*, 32(31), 10627–10636.
- Hafed, Z. M., Goffart, L., & Krauzlis, R. J. (2009). A neural mechanism for microsaccade generation in the primate superior colliculus. *Science*, 323(5916), 940–943.
- Hicheur, H., Zozor, S., Campagne, A., & Chauvin, A. (2013). Microsaccades are modulated by both attentional demands of a visual discrimination task and background noise. *Journal of vision*, 13(13), 18–18.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature neuroscience*, 3(3), 284–291.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1), 106–154.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3), 194–203.
- Itti, L., Koch, C., Niebur, E., et al. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11), 1254–1259.
- James, W. (1890). *The principles of psychology*. New York: Dover (1957).
- Kietzmann, T. C., Geuter, S., & König, P. (2011). Overt visual attention as a causal factor of perceptual awareness. *PLoS one*, 6(7), e22614.
- Ko, H.-k., Poletti, M., & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature neuroscience*, 13(12), 1549–1553.

- Kopecz, K., & Schöner, G. (1995). Saccadic motor planning by integrating visual information and pre-information on neural dynamic fields. *Biological cybernetics*, *73*(1), 49–60.
- Krauzlis, R. J., Goffart, L., & Haged, Z. M. (2017). Neuronal control of fixation and fixational eye movements. *Philosophical Transactions of the Royal Society*, *372*(1718), 20160205.
- Lefort, M., Boniface, Y., & Girau, B. (2011). Coupling bcm and neural fields for the emergence of self-organization consensus. In *From brains to systems* (pp. 41–56). Springer.
- Maggiani, L., Bourrasset, C., Quinton, J.-C., Berry, F., & Sérot, J. (2016). Bio-inspired heterogeneous architecture for real-time pedestrian detection applications. *Journal of Real-Time Image Processing*, 1–14.
- Milton, J., Insperger, T., & Stepan, G. (2015). Human balance control: dead zones, intermittency, and micro-chaos. In *Mathematical approaches to biological systems* (pp. 1–28). Springer.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24*(05), 939–973.
- Otero-Millan, J., Troncoso, X. G., Macknik, S. L., Serrano-Pedraza, I., & Martinez-Conde, S. (2008). Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator. *Journal of Vision*, *8*(14), 21–21.
- Pezzulo, G. (2008). Coordinating with the future: the anticipatory nature of representation. *Minds and Machines*, *18*(2), 179–225.
- Piaget, J. (1967). *The child's conception of space*. Norton.
- Poincaré, H. (1914). *Science and Method*. London: Thomas Nelson and Sons.
- Quinet, J., & Goffart, L. (2015). Does the brain extrapolate the position of a transient moving target? *The Journal of Neuroscience*, *35*(34), 11780–11790.
- Quinton, J.-C. (2010). Exploring and optimizing dynamic neural fields parameters using genetic algorithms. In *International Joint Conference on Neural Networks (IJCNN)*.
- Quinton, J.-C., Catenacci, N., Barca, L., & Pezzulo, G. (2014). The cat is on the mat. Or is it a dog? dynamic competition in perceptual decision making. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, *44*, 539–551. doi:
- Quinton, J.-C., & Girau, B. (2011). Predictive neural fields for improved tracking and attentional properties. In *International Joint Conference on Neural Networks (IJCNN)* (pp. 1629–1636).
- Quinton, J.-C., & Girau, B. (2012). Spatiotemporal pattern discrimination using predictive dynamic neural fields. In *Twenty first annual computational neuroscience meeting (CNS)* (Vol. 13, p. 1). BioMed Central.
- Rolfs, M. (2009). Microsaccades: small steps on a long way. *Vision research*, *49*(20), 2415–2441.
- Rolfs, M. (2015). Attention in active vision: A perspective on perceptual continuity across saccades. *Perception*, *44*(8-9), 900–919.
- Rosenbaum, D. A., Chapman, K. M., Weigelt, M., Weiss, D. J., & van der Wel, R. (2012). Cognition, action, and object manipulation. *Psychological bulletin*, *138*(5), 924.
- Rougier, N. P., & Vitay, J. (2006). Emergence of attention within a neural population. *Neural Netw.*, *19*(5), 573–581.
- Sandamirskaya, Y. (2014). Dynamic neural fields as a step toward cognitive neuromorphic architectures. *Frontiers in Neuroscience*, *7*, 276.
- Sandamirskaya, Y., & Schöner, G. (2010). An embodied account of serial order: How instabilities drive sequence generation. *Neural Networks*, *23*(10), 1164–1179.
- Schneegans, S., Spencer, J. P., Schöner, G., Hwang, S., & Hollingworth, A. (2014). Dynamic interactions between visual working memory and saccade target selection. *Journal of vision*, *14*(11), 9–9.
- Schöner, G. (2008). Dynamical systems approaches to cognition. *Cambridge handbook of computational cognitive modeling*, 101–126.
- Schöner, G., & Spencer, J. (2015). *Dynamic thinking: A primer on dynamic field theory*. Oxford University Press.
- Taouali, W., Goffart, L., Alexandre, F., & Rougier, N. P. (2015). A parsimonious computational model of visual target position encoding in the superior colliculus. *Biological cybernetics*, *109*(4-5), 549–559.



- Taylor, J. (1999). Neural bubble dynamics in two dimensions: Foundations. *Biological Cybernetics*, 80, 5167-5174.
- Uexküll, J. v. (1909). *The environment and inner world of animals*. Berlin.
- Viollet, S. (2014). Vibrating makes for better seeing: from the fly's micro-eye movements to hyperacute visual sensors. *Frontiers in bioengineering and biotechnology*, 2.
- Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2), 55-80.
- Zhang, K., & Sejnowski, T. J. (1999). Neuronal tuning: To sharpen or broaden? *Neural Computation*, 11(1), 75-84.