Journal of Biological Physics (2022) 48:195–213
https://doi.org/10.1007/s10867-021-09601-9

ORIGINAL PAPER

A neurocomputational model for the processing of conflicting information in context-dependent decision tasks

Francisco M. López[1] · Andrés Pomi[2]

Received: 14 September 2021 / Accepted: 24 December 2021 / Published online: 8 March 2022

Abstract

Context-dependent computation is a relevant characteristic of neural systems, endowing them with the capacity of adaptively modifying behavioral responses and flexibly discriminating between relevant and irrelevant information in a stimulus. This ability is particularly highlighted in solving conflicting tasks. A long-standing problem in computational neuroscience, flexible routing of information, is also closely linked with the ability to perform context-dependent associations. Here we present an extension of a context-dependent associative memory model to achieve context-dependent decision-making in the presence of conflicting and noisy multi-attribute stimuli. In these models, the input vectors are multiplied by context vectors via the Kronecker tensor product. To outfit the model with a noisy dynamic, we embedded the context-dependent associative memory in a leaky competing accumulator model, and, finally, we proved the power of the model in the reproduction of a behavioral experiment with monkeys in a context-dependent conflicting decision-making task. At the end, we discuss the neural feasibility of the tensor product and made the suggestive observation that the capacities of tensor context models are surprisingly in alignment with the more recent experimental findings about functional flexibility at different levels of brain organization.

Keywords Tensor representation · Context-dependent decision tasks · Filtering irrelevant information · Multi-attributes

1 Introduction

Context-dependent computation is a fundamental ability of our nervous system: a key stimulus is processed in different ways depending on other inputs that act as context. Obviously, considered in the abstract, the different simultaneous activations of large

---

✉ Andrés Pomi
pomi@fcien.edu.uy

[1] Interdisciplinary Center in Cognition for Education and Learning, Universidad de la República, José Enrique Rodó 1839 bis, 11200 Montevideo, Uruguay

[2] Group of Cognitive Systems Modeling, Biophysics and Systems Biology Section, Facultad de Ciencias, Universidad de la República, Iguá 4225, 11400 Montevideo, Uruguay

neuronal groups located in different sectors of the brain can be seen as equivalent components of a single composition of activity that defines the “state” of the system. But within a behavioral situation, whether in a natural setting or in an experimental task, the goal to be achieved by the individual constitutes the main stimulus, and the rest of the afferences, sensory, sensitive, emotional, etc. make up a “context” in which the subject is going to execute the action, make a decision, or interpret a percept.

This capacity of displaying adaptive behaviors was a fundamental evolutive acquisition of living species, drastically increasing their ability to survive. Already in invertebrates, different modes of circuit operation can be selected from the same circuit pattern generation by external influence [1]. Recently, state-dependent changes in functional connectivity have been found at different levels of brain organization, from circuit interactions in shaping population encoding [2] to the level of large-scale networks across cortical areas [3].

The relevance of context-dependent decisions is particularly highlighted in conflicting tasks, where the presence of misleading information is known to disrupt performance. The filtering of this misleading or irrelevant information is a permanent cognitive activity, closely related to attentional processes [4]. Sensory stimuli in ecological settings are usually multimodal. Nonetheless our brain can “gate in” visual information and filter out auditory stimuli while reading a book in a noisy room. Furthermore, we often find ourselves in situations where we receive stimuli from various sources of the same sensory modality. Besides being a remarkable capacity of the brain, the identification and separation of sources is an important technological challenge, which has been the subject of continuous advances in the field of signal engineering [5].

In experimental settings, this ability to filter out distracting information is particularly evident in perceptual decision tasks with noisy stimuli, such as the classical random dot kinetograms [6, 7]. In this common experimental paradigm used in many studies, subjects are presented with an animated cloud of apparently randomly moving dots, although a controlled fraction of these move coherently. The task consists of indicating the net direction of motion, with humans usually having to press a button while monkeys perform saccades to a target. The addition of conflicting information, e.g., colored dots, leads to lower performance in incongruent trials [8].

In this work, we propose a neurocomputational model for filtering irrelevant information in conflicting tasks, based on context-dependent memory matrices [9, 10]. These are vector state models that use tensor product variable binding [11], which demonstrated representational power and the possibility to develop theoretical neurocognitive approaches using the well-known capacities of linear algebra [12, 13]. These models share some of their key features, such as the tensor product variable binding and high-dimensional distributed vector representations with the paradigms of vector symbolic architectures and hyperdimensional computing [14–16].

Our main idea is to capture the representation of different sensory stimuli composed as a tensor product and then embed this core structure in the dynamical noisy scheme of a sequential sampling model. The model captures the main phenomenology of the random dot kinetogram context-dependent decision task in the experiment of Mante et al. [8]. It should be noted that our analysis has been a qualitative one. We merely showed that the model put forward is capable of explaining the behavior observed in monkeys for a certain set of parameters and neural codes. The aim of the model is to contribute to the ongoing modeling exploration over the multiple and important advances that have recently occurred in this area and in no way can be considered a definitive statement of how these cognitive processes arise from the neural substrate.

The article is organized as follows. In the next section, we present context-dependent associative memory models. In Sect. 3, we generalize the model to make it capable of processing inputs composed of multiple features and cue-dependent different tasks. In Sect. 4, we complete our model by embedding it in the frame of the classical leaky competing accumulator model, in order to endow it with a noisy dynamic, and then, in Sect. 5, we prove the potential of the model by reproducing the behavioral data from a cue-dependent task. We close with a discussion and declaration of envisioned future work.

2 A review of context-dependent associations in matrix memory models

Distributed vector coding and associative matrix memory models were born in the neurosciences at the beginning of the seventies [17], inspired by the content-addressable nature of information storage in the brain and by the suggestive observations of Longuet-Higgins [18] and Dennis Gabor [19]. These systems provide a distributed mapping from high-dimensional input vectors to high-dimensional output vectors. The vector representation relies on the fact that any kind of information in the brain is coded by the activity of large groups of neurons, as we now certainly know [20]. High-dimensional vectors are, therefore, natural mathematical variables to represent neural activity.

Cooper [21], deeply impressed by these models, turned to the field of theoretical neurobiology immediately after obtaining the Nobel Prize for his contributions in superconductivity and made decisive contributions to the field [21, 22]. These models, found in an independent way by Anderson [23] and Kohonen [24], were generated in a constructive way, parsimoniously incorporating neurobiological knowledge that is still in force. This is particularly revealed in [25].

A pattern of activation over a group of n neurons is represented by a vector $\mathbf{x} \in R^{\mathrm{n}}$. The number of neurons involved in a modality-specific cognitive unit should be considered large, even though activity may be sparse. Each neuron is represented in a fixed position of the inputs of a vector by a measure of its activity, usually its mean firing rate [26].

Assuming the simplest version of a Hebbian synapse, in the projection of a group of afferent axons F onto the dendritic trees of another group of neurons G, an associative matrix memory $\mathbf{M}$ is naturally given by the sum of the outer products of each pair of associated input–output activity vectors:

$$\mathbf{M} = \sum_{i=1}^{k} \eta_i \mathbf{g}_i \mathbf{f}_i^T \tag{1}$$

where $\{\mathbf{f}_i\}$ are the input vectors and $\{\mathbf{g}_i\}$ the output vectors. The coefficients η_i indicate the relative strength of each stored association. They represent the different frequency of presentation of the pairs in the learning history. In the presence of a new input $\mathbf{f}$, retrieval from the associative memory is performed by

$$\mathbf{Mf} = \sum_{i} \eta_i \mathbf{g}_i \langle \mathbf{f}_i, \mathbf{f} \rangle \tag{2}$$

where the inner products $\langle \mathbf{f}_i, \mathbf{f} \rangle$ give a measure of similarity between the presented input and all previously stored inputs. Thus, if the presented input was stored in the memory and it is orthogonal to all other stored inputs, the associated output is perfectly recovered.

Although the orthogonality of the vector patterns $\{\mathbf{f}_i\}$ seems at first glance to be a very demanding condition, several circumstances normally present in nervous systems make it possible to ensure an approximate orthogonality of the vectors that encode different concepts. In fact, the high dimensionality of the vectors favors that the scalar products between pairs of them have values close to zero if the code is sparse, and the same happens for random codes under many symmetrical statistical distributions of their scalar components centered around zero [17, 27]. Also note that this *quasi-orthogonality* can emerge from a spatially patched activity distribution, like the one seen on topographic maps. A regionalized activity in a neural layer can be represented in vector form, where the different spatial zones are mapped in different blocks of the vectors, generating scalar products close to zero for vectors that encode the activation of different zones. For this reason, the existence of topographic maps can be seen, from a computational point of view, as a strategy of the nervous systems to orthogonalize its inputs.

Otherwise, some degree of overlap with other inputs may result in interference between outputs. This flawed retrieval may be undesired for artificial agents and digital computers, but it is a good model of the brain. Besides, the distributed nature of the information provides high reliability in the face of random losses of components both at the level of vectors and matrices [27]. These memories also present the ability to generalize from the learning of examples, as it was early recognized [22].

But these classic memories suffer from a significant disability in being unable to solve the bifurcation problem, as Anderson called it [28]: the possibility of triggering different associative chains, given the same present state. That is, given the same input, an Anderson-Kohonen associative memory always responds with the same output. This limitation is not solved by lengthening the input vector with a sector that represents a context, since in this type of "additive contextualization," the outputs are not separable due to the linear character of the operator. This is related to the impossibility of these early models to compute the XOR logical gate, a flaw that they share with one-layer perceptrons, as Minsky and Papert famously pointed out [29].

To overcome this inability, two different lines of neural models were developed. The first of them is the famous multilayer perceptron trained with backpropagation [30], the ancestor of today's multilayer networks that are revolutionizing artificial intelligence. Although computationally performant, this model lacks biological plausibility. Therefore, although it is the model on which most applications of artificial neural networks were based by far, it is not an attractive model for the relatively small community of neurobiologically oriented model developers.

The other proposal is based on a preprocessing of the stimulus vector with another vector acting as a multiplicative context [9]. "Context-dependent associative memories" have two virtues for theoretical neurobiologists: they rely on feasible neurobiological mechanisms (we will discuss this topic later in Sect. 6), and because they rest on matrix algebra, they allow the development of mathematical theoretical approaches for complex cognitive activities [31], whereas it is impossible to do algebra from the backpropagation model. Almost simultaneously and independently, Smolensky found the same solution [11], although he devoted his later developments to exploit an interpretation of the model as a filler-role binding, especially suitable for linguistic processing [32].

In context-dependent associative memory models, the main stimulus and the inputs acting as context are represented by vectors that are premultiplied by the Kronecker product. Representing the patterns of activity of neuronal context afferences with vectors $\{\mathbf{p}\}$, a context-dependent associative memory is given by the equation

$$\mathbf{M} = \sum_i \sum_j \eta_{ij}\,\mathbf{g}_{ij}(\mathbf{p}_j \otimes \mathbf{f}_i)^T \tag{3}$$

where for each input $\mathbf{f}_i$, different outputs $\mathbf{g}_{ij}$ are associated depending on the accompanying context $\mathbf{p}_j$. The subscript j runs through the entire set of the possible contexts given the kind of inputs considered. A same context may be present with different inputs and when a given context $\mathbf{p}_j$ never appeared with an input $\mathbf{f}_i$, the corresponding parameter η_{ij} will be zero.

The symbol ⊗ means the Kronecker product, a particular case of the tensor product of two matrices of arbitrary dimensions [33]. For two matrices **A** and **B**, their Kronecker product is $\mathbf{A} \otimes \mathbf{B} = [a_{ij}\,\mathbf{B}]$. The fundamental properties of this product are revised in Appendix 1.

Presenting to the memory a new input **f** in a context **p** yields.

$$\mathbf{M}(\mathbf{p} \otimes \mathbf{f}) = \sum_{i,j} \eta_{ij}\,\mathbf{g}_{ij}(\mathbf{p}_j^T\mathbf{p})(\mathbf{f}_i^T\mathbf{f}) = \sum_{i,j} \eta_{ij}\,\mathbf{g}_{ij}\big\langle \mathbf{p}_j, \mathbf{p}\big\rangle\langle \mathbf{f}_i, \mathbf{f}\rangle \tag{4}$$

It can be observed that the retrieved output of the memory is a linear combination of all the stored outputs. Each output vector $\mathbf{g}_{ij}$ in the memory is weighted by two scalar products: one compares the similarity of the input with all the stored inputs, and the other evaluates the similarity of the present context with all the contexts stored in the memory. Orthogonal vectors cancel out the corresponding dot product, giving the double filter a conjunctive character. As before, if the inputs and contexts form quasi-orthogonal sets, this model can achieve perfect recall.

These non-linear vector state models enable the capture of complex cognitive functions such as the selective extraction of features from complex patterns [34], adaptive searching [12] or logical reasoning [35, 36]. The thematic packaging of information, an essential feature of context-dependent associative memories, shows a kinship between these models and the vector space models used to extract information from databases as it is the case with latent semantics [12] and with search engines [10]. Other applications of this formalism include the organization of thematically ordered sequences in the production of language [37], the exploration of the cognitive processes involved in medical diagnosis [38], and the possibility of spatial organizations of memories in semantic topographies [39]. With the conjunction of the vector representation of cognitive states, the ability for pattern association based on simple Hebbian synapses, and the adaptive capabilities provided by tensor product preprocessing, these models are a strong candidate for developing a universal theory of cognition. Furthermore, their vector–matrix algebraic nature opens the possibility of developing a unified framework for theory advancement in cognitive neurosciences.

To apply the model to context-dependent decision tasks in which the system has to respond to stimuli with multiple attributes (eventually conflicting), we must generalize it, which we will do in the next section, and then it will be necessary to equip our model with a dynamic, to which we will dedicate ourselves in Sect. 4.

3 Generalization to cue-dependent decisions for stimuli with multiple attributes

In this section, we reformulate the model to make it capable of processing inputs composed of multiple features and cue-dependent different tasks, motivated by the behavioral data described in [8] that we are going to simulate in Sect. 5.

In this particular case, we are faced with a situation in which there is no stimulus and its attributes, but the stimulus is the set of attributes. The concatenation of attributes in a single extended stimulus vector could be a naive first way of attacking the problem. However, it is well known that this additive composition of vectors prevents the separation of the outputs, since the coincident sectors of the input vectors generate non-zero scalar products when processing the inputs. This situation could be overcome by the multiplicative composition of the inputs through the tensor product.

Our first step is to format context-dependent associative memories to model decisions in the presence of stimuli with multiple attributes. Let us consider an experiment where participants are shown stimuli with multiple features and instructed to perform specific responses depending on those features. Until now, we had presented associative memories in an abstract and general way, for which we had used the classical nomenclature, naming the input vectors with **f** and the output vectors with **g**. To mark that it is now a specific representation, carried out in some sectors in the nervous system, we change the nomenclature, and we will use **s** for stimuli and **z** for response vectors.

The n independent features or attributes of the stimulus are represented by vectors $\mathbf{s}^1$, $\mathbf{s}^2$,…., $\mathbf{s}^n$. Without loss of generality and for the sake of mathematical simplicity, we consider that all these features are represented by vectors of the same dimension m and also that the set of potential attributes and their arbitrary order is the same for all the stimuli of our experimental situation. In neurobiology, the different attributes of a certain type of stimulus, such as those that arise from a fixed experimental setting, surely come from the cortical areas that are activated by different characteristics or modalities of the stimulus. So, we assume that the stimuli are well structured by that biological domain, and in this article, we do not deal with the mechanisms that could create those structured representations.

A stimulus (previously an input denoted by **f**) is now a structured object consisting of multiple attributes encoded in a single vector $\mathbf{s}_\mathrm{i}$ of dimension $(m^\mathrm{n} \times 1)$ formed by the multiple Kronecker product of its attributes:

$$\mathbf{s}_i = \mathbf{s}^1 \otimes \mathbf{s}^2 \otimes \cdots \otimes \mathbf{s}^n = \bigotimes_{k=1}^{n} \mathbf{s}^k \tag{5}$$

A memory that stores associations between stimuli $\mathbf{s}_\mathrm{i}$ and their responses $\mathbf{z}_\mathrm{i}$ is given by

$$\mathbf{M} = \sum_i \eta_i \mathbf{z}_i \mathbf{s}_i^T = \sum_i \eta_i \mathbf{z}_i (\bigotimes_k \mathbf{s}_i^k)^T \tag{6}$$

During an experimental trial, a multi-attribute stimulus (a) is presented to the subject. The corresponding vector $\mathbf{s}_a = \bigotimes_k \mathbf{s}_a^k$ encodes the neural activation produced by that stimulus across all attributes. The product of the associative memory **M** and the stimulus vector $\mathbf{s}_\mathrm{a}$ is given by

$$\mathbf{M}\mathbf{s}_a = \sum_i \eta_i \mathbf{z}_i (\bigotimes_k \mathbf{s}_i^k)^T (\bigotimes_k \mathbf{s}_a^k) \tag{7}$$

and from the properties of the Kronecker product (see Appendix 1), it can be seen that the product of the vectors is the scalar product of their pairwise inner products corresponding to the n attributes:

$$\mathbf{M}\mathbf{s}_a = \sum_i \mathbf{z}_i \eta_i \prod_{k=1}^{n} \left\langle \mathbf{s}_i^k, \mathbf{s}_a^k \right\rangle \tag{8}$$

Hence, each attribute is compared individually to the corresponding attributes of the stored stimuli. In the absence of difference at the level of η_i, the preponderant output $\mathbf{z}_i$ in the resulting linear combination is that associated with a similar stimuli, calculated by means of an inner product.

Redefining $\sigma_{i,a}^k = \left\langle \mathbf{s}_i^k, \mathbf{s}_a^k \right\rangle$, the previous equation can be rewritten as:

$$\mathbf{M}\mathbf{s}_a = \sum_i \mathbf{z}_i \left[\eta_i \prod_{k=1}^{n} \sigma_{i,a}^k \right] \tag{9}$$

which shows how attribute-level similarities are then multiplied to obtain the final weights (between brackets) in the system response. Thus, context-dependent associative memories can store and recall mappings from multi-attribute stimuli to the corresponding responses.

Now, as a second step, we expand this model to accomplish different tasks depending on given cues that we incorporate as multiplicative contexts in our memory. A stimulus vector $\mathbf{s}_a$ accompanied by a cue context vector $\mathbf{c}$ can be expressed by a tensor composition as follows:

$$\mathbf{c} \otimes \mathbf{s}_a = \mathbf{c} \otimes \ [\bigotimes_{k=1}^{n} \ \mathbf{s}_a^k] \tag{10}$$

As we have discussed in the Introduction, the distinction between stimulus and context is, in general, an arbitrary decision. As can be seen in Eq. (10), the cue context vector $\mathbf{c}$ could be treated as a first entry within the set $\{\mathbf{s}^{\mathrm{k}}\}$. Here we assume that the cue context is, for example, an order transmitted to the experimental subject by verbal or visual means and that the composition of the different stimuli is shaped by the wiring of the nervous system.

Thus, a generalized context-dependent memory matrix taking different cues into account can now be written. This global memory is given by

$$\mathbf{M} = \sum_{i,j} \eta_{ij} \mathbf{z}_{ij} (\mathbf{c_j} \otimes \mathbf{s}_i)^T = \sum_{i,j} \eta_{i,j} \mathbf{z}_{i,j} (\mathbf{c}_j^T \otimes \ \ [\bigotimes_k \mathbf{s}_i^k]^T) \tag{11}$$

where the double sum spans over all stimuli and cues. The strengths of the associations η_{ij} and the output vectors $\mathbf{z}_{ij}$ depend on both the stimulus and the cue, because the same stimulus can be associated with different responses under different contexts.

Assume that cue context vectors form an orthonormal set, such that $\langle \mathbf{c}_h, \mathbf{c}_{h'} \rangle = \delta_{hh'}$. Then, presenting a cue $\mathbf{c}_v$ followed by the multi-attribute stimulus $\mathbf{s}_a = \bigotimes_k \mathbf{s}_a^k$ yields the output

$$\mathbf{M}(\mathbf{c}_v \otimes \mathbf{s}_a) = \sum_{i,j} \eta_{ij} \mathbf{z}_{ij} (\mathbf{c}_j^T \mathbf{c}_v) \otimes [(\bigotimes_k \mathbf{s}_i^k)^T (\bigotimes_k \mathbf{s}_a^k)] = \sum_i \eta_{iv} \mathbf{z}_{iv} \prod_{k=1}^{n} \sigma_{i,a}^k \tag{12}$$

which is a linear combination of the output vectors $\mathbf{z}$ instructed in memory. Here we assume that the nervous system ends up filtering the heaviest component in this output combination, which is expressed in Eq. (13). The process of highlighting this predominant component and attenuating the rest of the components could be implemented in the nervous system by some recursive mechanism, whose existence is presupposed, and we do not explore in this article. The selected response will be $\mathbf{z}_{*v}$ corresponding to

$$\eta_{*v}\prod_{k=1}^{n}\sigma_{*,a}^{k} > \eta_{iv}\prod_{k=1}^{n}\sigma_{i,a}^{k} \quad \forall\, i \neq * \tag{13}$$

Two interesting conclusions can be drawn from this equation. The first is that the parameters η are of crucial importance, since the stronger associations will be favored and more easily selected. Second, the overall product of all features is used to make a choice, so that no single feature drives the decision-making process on its own. Note that, however, due to the multiplicative character of the scalar products, a mismatch in only one feature cancels the contribution of the corresponding output z in the linear combination of output.

The model, as it is so far, only produces responses in one step, single associations, no matter how complex and structured the memory or input is. For a real decision-making task, we must have a model that captures the dynamics of the process, in particular the progressive accumulation of evidence and, furthermore, that is capable of producing errors, having a somewhat stochastic behavior. That is our assignment for the next section.

4 Embedding of the tensor model in a dynamic framework with noisy inputs

Psychologists often use response times and error rates in behavioral experiments as quantitative measures of the degree to which irrelevant information is filtered. In order to reproduce those results with associative memories, one needs to consider a dynamic evolution for the response selection process. In this section, we show how a leaky, competing accumulation of the matrix output can provide such dynamics.

The *leaky competing accumulator model* [40] (in what follows LCA) proposes that decision-making involves the accumulation of noisy information by an ensemble of accumulators, one for each alternative in a decision-making situation. Each accumulator is thought to correspond to a large population of neurons (in parietal or prefrontal cortex [41, 42]), in competition with the neurons in other populations. At any time, the *decision state* of the system corresponds to the pattern of activation across the ensemble of accumulators. The process continues until one of the accumulators reaches a threshold and the corresponding response becomes selected (Fig. 1). In the model, the state of each accumulator is represented with a single activation value, and its time evolution is described by the differential equation:

$$dx_i = \left(\mu_i - \alpha x_i - \beta \sum_{j\neq i} x_j\right)\frac{dt}{\tau} + \xi_i\sqrt{\frac{dt}{\tau}} \tag{14}$$

The term μ_i represents a constant external input to the accumulator. The parameters α and β corresponds to the magnitude of leak and competition, respectively. When leakage is high, old accumulated evidence is “forgotten” quickly, so that newer evidence plays a higher role in the decision-making process. The competition term generates coupling between the equations of the different accumulators. ξ_i represents zero-mean Gaussian noise. The term *dt* represents the time step, while τ is a characteristic time scale for integration. The accumulator is set to zero when its equation results in a negative value, making the model non-linear.

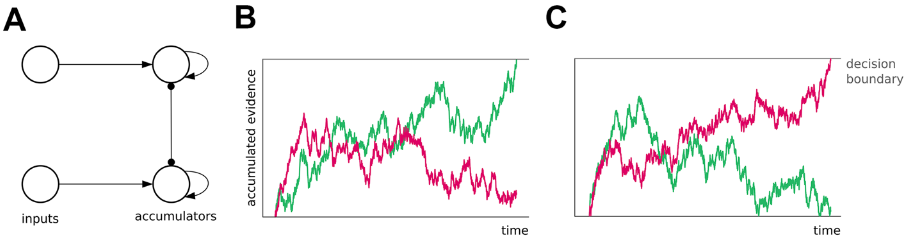

Fig. 1 Overview of the leaky competing accumulator model (LCA). **A** Architecture of the LCA model consisting of two input nodes and two accumulator nodes with leakage and lateral inhibition. **B**–**C** Examples of the time evolution of the accumulated evidence for correct and incorrect trials in the LCA model. Accumulation finishes once one of the variables reaches the decision boundary

In the simplest implementation of the LCA model, the accumulators are initially set to zero and evolve according to (14). In a task under standard conditions, the selected response corresponds to the accumulator that first reaches a threshold boundary (see Fig. 1). Alternatively, time-controlled tasks force subjects to make a decision when a certain time has elapsed, so that the highest accumulator is the one chosen. For a detailed analysis of the dynamics of LCA and the interplay between leak and competition, see [43].

Let us consider the simple case of two stimuli, a and b, such that when stimulus a is shown, the expected response corresponds to accumulator x_1 and when stimulus b is shown, it corresponds to x_2. Subjects may occasionally make mistakes, which is captured by the inclusion of step-by-step noise. Assuming that there are no further difficulties involved in the task, the following conditions should be met: μ_1 (a) $>$ μ_2 (a) and μ_1 (b) $<$ μ_2 (b). Usher and McClelland [40] suggest an extra constraint: that the total information distributed over the accumulators should be kept constant for all stimuli. This is done by imposing $\mu_1+\mu_2=1$. Note that the LCA model is indifferent to the representation of stimuli in possible previous layers and, due to its phenomenological nature, does not make assumptions about the representation of stimuli in the brain either. In the remainder of this section, we show how a vector representation of the neural variables and the use of linear associative memories allow the explicit inclusion of stimuli within the LCA model at the same time that it provides the desired noisy dynamics to our tensor contexts associative memory model.

Let us define the column vectors $\mathbf{x} = \left[x_1, x_2\right]^T$, $d\mathbf{x} = \left[dx_1, dx_2\right]^T$, $\boldsymbol{\mu} = \left[\mu_1, \mu_2\right]^T$, $\boldsymbol{\xi} = \left[\xi_1, \xi_2\right]^T$. It can be readily seen that a system of two coupled differential equations as (14) can be rewritten as a vector–matrix equation:

$$d\mathbf{x} = \left[\boldsymbol{\mu} - \mathbf{D}\mathbf{x}\right]\frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{15}$$

where **D** is a bisymmetric square matrix with positive entries, which we will call the decay matrix, given by $\mathbf{D} = \begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}$. Equation (15) holds for the general LCA model with m multiple accumulators, where vectors **x**, d**x**, **μ**, and **ξ** are m-dimensional and the decay matrix **D** has dimensions mxm with values α in the main diagonal and β elsewhere. The term (**μ**—**Dx**) indicates the noise-free dynamics. Initially, with $\mathbf{x}=\mathbf{0}$, accumulation is proportional to input **μ**. As **x** changes so does d**x**. Since **D** is bisymmetric, if $\alpha \neq \beta$, then its determinant is non-zero, and there is a unique equilibrium $\mathbf{x}_{\text{eq}}$ such that $\boldsymbol{\mu}=\mathbf{D}\mathbf{x}_{\text{eq}}$ which, if $\alpha>\beta$, is an attractor point; otherwise it is a repeller. Alternatively, if $\alpha=\beta$, the determinant is zero, and there are either infinite equilibrium points if evidence for all accumulators is the same or there are no equilibrium points.

In what follows, we will consider that the input to the accumulators, $\boldsymbol{\mu}$, is the output of an associative memory matrix **M** that accomplishes the recognition and the transformation of stimuli **s** into their associated readouts: $\boldsymbol{\mu} = \mathbf{Ms}$. The representation **s** describes the neural activity associated with the encoding of the sensory input.

In general, while **D** remains a square matrix, $\mathbf{M}_{(mxm')}$ can be rectangular, since it stores a mapping from an m'-dimensional stimuli space to the m-dimensional accumulators' space, for m-dimensional choice ($m' \geq m$). For the moment, we will assume that **M** is given and fixed during an experiment.

The same way as each accumulator is considered to capture average firing rates of entire populations of parietal or prefrontal neurons, we suggest that stimuli vectors represent distributed firing rates of large groups of neurons from modality- and attribute-specific brain regions [41]. It is reasonable then to assume that m' can be as large as necessary, although, mathematically, a dimension $m' = m$ is sufficient for adequate discrimination.

Considering multi-attribute stimuli vectors **s** as those in (5) and equipped with a matrix memory like the one in (6), we have an extension of classical LCA which we call *multi-attribute leaky competing accumulator* (MLCA) model:

$$d\mathbf{x} = \left[\mathbf{M}\left(\bigotimes_{k=1}^{n} \mathbf{s}^k\right) - \mathbf{Dx}\right]\frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{16}$$

As before, the decay matrix **D** is a bisymmetric square matrix with leakage terms α in the main diagonal and lateral inhibition terms β elsewhere.

Hence, given a particular multi-attribute stimulus $\mathbf{s}_\mathrm{a}$, $\mathbf{Ms}_\mathrm{a}$ is given by Eqs. 7–9, and the dynamics of the accumulators will be governed by

$$d\mathbf{x} = \left[\sum_i \mathbf{z}_i\left(\eta_i \prod_{k=1}^{n} \sigma_{i,a}^k\right) - \mathbf{Dx}\right]\frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{17}$$

On average, the accumulator that most frequently wins the race when stimulus a is presented is the one corresponding to $\mathbf{z}_\omega$ such that the accumulator-level evidence is maximized. The index ω satisfies

$$\eta_\omega \prod_{k=1}^{n} \sigma_{\omega a}^k > \eta_i \prod_{k=1}^{n} \sigma_{ia}^k \quad \forall\, i \neq \omega \tag{18}$$

which can be thought of as a recall function. When stimulus (a) is presented, the evidence for all accumulators is calculated based on the stimuli stored in the associative memory. The accumulator level evidence will be highest for the alternative that maximizes the entire product, even if other alternatives have higher attribute-level evidences for certain attributes. Importantly, the remaining alternatives play an important role in the dynamics of the MLCA model due to lateral inhibition and the stochastic nature of the accumulation.

The final step needed to complete our construction is to incorporate to the model the capability of focusing attention to different features of the stimuli, depending on a given cue. The stimulus accompanied by its context vector **c** can be expressed by a tensor composition as follows: $\mathbf{c} \otimes \mathbf{s}_a$ (10). With a memory matrix like the one in (11), our final MLCA model governs the dynamics according to

$$d\mathbf{x} = \left[\mathbf{M}\left(\mathbf{c} \otimes \mathbf{s}_a\right) - \mathbf{D}\mathbf{x}\right]\frac{dt}{\tau} + \xi\sqrt{\frac{dt}{\tau}} \tag{19}$$

Now, we have developed the tool for studying the dynamics of context-dependent decisions with conflicting attributes. In the next section, we apply our model to the discussion of an example from a two-alternative task with two-attribute stimuli. Given the experimental data, we will see that the MLCA model not only allows for an understanding of the dynamics of evidence integration but also allows for the explicit study of the representation codes of stimuli.

5 Reproducing the behavioral data from a cue-dependent task

In this section, we test the capabilities of the model to reproduce the behavioral data of Mante et al. [8], who tested two macaque monkeys on a color and motion random dot kinetogram.

The monkeys A and F were trained to fixate a screen and wait for a contextual cue that indicated the relevant attribute for each trial: a yellow square indicated that the monkeys were to respond based on the motion direction of the dots (right or left), whereas a blue cross indicated that responses should be based on their color (green or red) (see Fig. 2). The dots were presented for 750 ms; during that time, the monkeys integrated the noisy sensory inputs towards a choice. Following a delay of 300 to 1500 ms, the monkeys were to make a saccade in the direction of their chosen response. The experimenters varied the motion and color coherences in both contexts and studied the monkeys' psychophysical performances by plotting psychometric curves. During the experiment, they recorded nearly 1500 units in the frontal eye field, a prefrontal area responsible for the selection and preparation of saccades [44].

Mante et al. [8] suggest that stochastic models of selection and integration [41, 45] are unable to account for context-dependent behavior. Instead, the authors use a recurrent neural

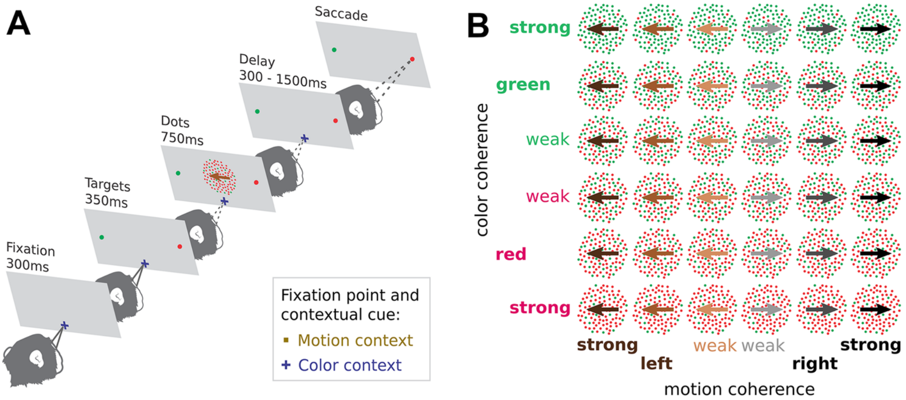

Fig. 2 Details of the behavioral task from Mante et al. [8]. Adapted by permission from Springer Nature: Mante, V. et al. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84 (2013), Fig. 1. **A** Timeline. Monkeys were trained to fixate a contextual cue which indicated whether motion or color were to be attended. A random dot kinetogram with preset motion and color coherences was shown for 750 ms; after a random delay, the fixation point disappeared, and the monkeys performed a saccade to the selected target. **B** Stimuli. Motion coherences were set to 5, 15, and 50% moving in the defined direction, with the remaining dots moving randomly; color coherences were set to 6, 18, and 50% in the defined color, with the remaining dots colored randomly

network which receives color and motion sensory evidence as inputs, as well as a binary contextual cue. The network outputs a single value, integrated over time, corresponding to its choice. This recurrent neural network qualitatively reproduces the population trajectories of the recorded units in the subspace of motion, color, and choice. Nonetheless, the model fails to reproduce the psychometric curves for the irrelevant attributes, especially for monkey F who consistently failed to ignore the irrelevant attribute (see Fig. 4). Next, we show how the MLCA model presented in this paper can correctly account for the psychophysical behavior without contradicting the single-unit recordings of [8]. Note that given the abstract nature of our model, the parameter values were set to fit the experimental data. It is not within the scope of this work to provide insights on the possible in vivo mechanisms by which these values are established.

Given a context-dependent associative memory **M** that stores the associations of the cue-dependent relevant attributes with their corresponding responses, and a decay matrix **D**, the dynamics of the MLCA model is given by

$$d\mathbf{x} = \left[\mathbf{M}(\mathbf{c}_{cue} \otimes \mathbf{p}_{motion} \otimes \mathbf{q}_{color}) - \mathbf{D}\mathbf{x}\right]\frac{dt}{\tau} + \xi\sqrt{\frac{dt}{\tau}} \tag{20}$$

where $\mathbf{c}_{cue} \in \left\{\mathbf{c}_{motion}, \mathbf{c}_{color}\right\}$ is the contextual cue, either for motion or for color (the two possible relevant attributes in this task), while $\mathbf{p}_{motion} \in \left\{\mathbf{p}_{100\%\ left}, \cdots, \mathbf{p}_{100\%\ right}\right\}$ and $\mathbf{q}_{color} \in \left\{\mathbf{q}_{100\%\ green}, \cdots, \mathbf{q}_{100\%\ red}\right\}$ are the motion coherence and the color coherence of the stimuli, respectively. For example, a trial with *50% left motion* coherence and *25% green color* coherence where the monkey should attend to motion is encoded by the input vector $(\mathbf{c}_{motion} \otimes \mathbf{p}_{50\%\ left} \otimes \mathbf{q}_{25\%\ green})$. That is, with a motion cue, left-moving dots should lead to a left saccade regardless of their color, whereas with a color cue, green dots should lead to a left saccade regardless of their motion. Furthermore, given the asymmetry of these associations, left and green dots in both contexts should have stronger leftward evidence accumulation than left and red in the motion context or right and green in the color context.

Please note that although the dimensionality of neural groups is always high, for computational convenience, we use a low dimensional vector representation. This is acceptable because in this task, the important thing is the recognition of the patterns involved, which is mediated by the angle between the vectors and is not influenced by their dimensions or the code used.

Each of the motion and color vectors is expected to encode a scalar coherence, and similar coherences must have similar vector representations with high inner products. The encodings will therefore follow a one-dimensional path between the representations for the two extreme coherences. This allows for a simplification by assuming that the matrix memory only stores the associations for the extreme coherences (i.e., the corners in Fig. 2B) with the corresponding responses. Defining $\mathbf{r}_{left}$ and $\mathbf{r}_{right}$ as the response codes to the left and to the right, respectively, and $\eta > \eta'$ as the scaling rates for the main diagonal (i.e., left-green and right-red) and anti-diagonal (i.e., left-red and right-green) stimuli, respectively, a possible context-dependent associative memory for the experiment of Mante et al. [8] would be

$$\begin{aligned}\mathbf{M} = \;&\eta\left(\mathbf{r}_{left}\mathbf{c}^T_{motion} + \mathbf{r}_{left}\mathbf{c}^T_{color}\right) \otimes \mathbf{p}^T_{left} \otimes \mathbf{q}^T_{green} + \\ &\eta'\left(\mathbf{r}_{left}\mathbf{c}^T_{motion} + \mathbf{r}_{right}\mathbf{c}^T_{color}\right) \otimes \mathbf{p}^T_{left} \otimes \mathbf{q}^T_{red} + \\ &\eta'\left(\mathbf{r}_{right}\mathbf{c}^T_{motion} + \mathbf{r}_{left}\mathbf{c}^T_{color}\right) \otimes \mathbf{p}^T_{right} \otimes \mathbf{q}^T_{green} + \\ &\eta\left(\mathbf{r}_{right}\mathbf{c}^T_{motion} + \mathbf{r}_{right}\mathbf{c}^T_{color}\right) \otimes \mathbf{p}^T_{right} \otimes \mathbf{q}^T_{red}\end{aligned} \tag{21}$$

where each line in (21) corresponds to one of the corners in Fig. 2B. It is assumed that the matrix memory stores abstractions of the stimuli the monkeys are exposed to during training. For example, the monkeys are trained to perform a rightward saccade upon seeing a majority of red dots in a color context, regardless of exactly how many red dots there are and how they are moving. This is a long process. For our model, we are not concerned with the establishment of this memory but rather with the fact that during the testing phase of the experiment, the monkeys have already learned these associations.

We consider 11 motion coherences and 11 color coherences and perform 1000 simulations for each combination with both possible cues. The parameters from the model are heuristically optimized such that the behavior of the model reproduces the behavior of the monkeys. When fitting the parameters, we consider a single set of parameters for both monkey across all experimental conditions, with the exception of the cue vectors that are orthogonal for monkey A but not for monkey F, thus capturing the latter's worse performance. The source code for the simulations is available under an MIT license at https://github.com/franmlopez/mlca. More details about the simulation parameters are provided in Appendix 2.

Let us call γ the inner product of the stored and the presented stimuli motion vectors and ν the inner product of the stored and the presented stimuli color vectors. Upon seeing a stimulus $(\mathbf{p}, \mathbf{q})$ in the motion context, the output of the memory is given by

$$\begin{aligned}\mathbf{M}(\mathbf{c}_{motion} \otimes \mathbf{p} \otimes \mathbf{q}) = \mathbf{r}_{left}\left[\eta\, \gamma_{left}\, \nu_{green} + \eta' \gamma_{left}\, \nu_{red}\right] + \\ \mathbf{r}_{right}\left[\eta\, \gamma_{right}\, \nu_{red} + \eta' \gamma_{right}\, \nu_{green}\right]\end{aligned} \tag{22}$$

The bracketed terms are the readouts and therefore set the evidence in favor of the left and right responses, respectively. It can be immediately seen that if the $\mathbf{p}$ motion coherence is towards the left, then $\gamma_{left} > \gamma_{right}$, and the left motion will be selected more frequently (see Fig. 3). Interestingly, the asymmetry imposed by setting $\eta > \eta'$ means that green color coherences in the motion context have higher evidence to the leftwards accumulator than red color coherences, as seen in the monkeys' behavioral data but not in the simulations of Mante et al. [8].

Figure 4 shows the psychometric curves obtained from the experimental data of monkeys A and F, as well as the MLCA simulations. Our model is capable of reproducing the behavioral results, particularly the influence of the irrelevant attributes. Changing the color and motion coherence levels affects performance in both color and motion contexts. According to Mante et al. monkey F was consistently worse than monkey A, as revealed by the slope of the psychometric curves when changing the irrelevant attribute. We reproduce this difference by using two different associative memories, with monkey F's stored vectors being less orthogonal but all other parameters set as equal. Thus, in our model, both monkeys receive the same input information, but monkey A is able to filter out the irrelevant attribute better at the level of the context-dependent associative memory.

These results show that explicit stimuli representations and their readouts by means of context-dependent associative memories are relevant mechanisms to take into account when modeling perceptual decision-making.

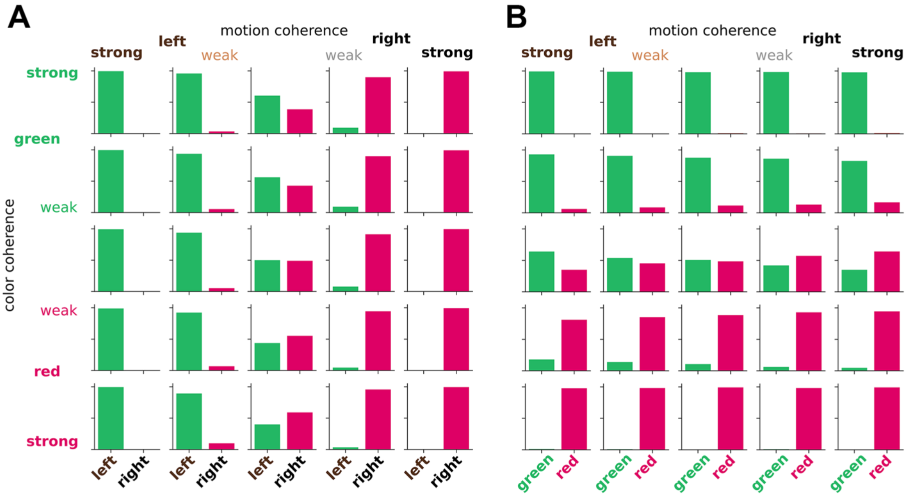

Fig. 3 Response histograms for simulations using the MLCA model. **A** Motion context; **B** color context. Each histogram corresponds to a specific motion coherence (horizontal axis) and color coherence (vertical axis), with bar heights indicating the percentage of correct responses. One thousand simulations were performed for each motion and color coherence in each context. Due to the stochastic integration, the same stimulus can entail any of the two possible responses. These results show a high dependency on the task-relevant attribute and a low dependency on the task-irrelevant attribute. However, the task-irrelevant attribute has a considerable effect on task-relevant neutral coherences (i.e., central column in **A**, central row in **B**). Mante et al. [8] do not use entirely neutral stimuli in their experiment

6 Discussion and future work

In this article, we have presented an extension of a vector state model that uses tensor product variable binding, context-dependent associative memories, to contribute a new tool to the modeling of context-dependent neurocomputations. This model has the advantage of accomplishing adaptive associations without losing on the way the matrix operators of linear algebra. Here, we modified the model to make it capable of performing dynamical decisions with noisy and progressive accumulation of evidence in cue-dependent tasks, where the stimuli have multiple and conflicting attributes and proved its potentialities reproducing the data and qualitative behavior in the experiment of [8].

In what follows, we will make some considerations about the neural feasibility of the tensor product. One of the requirements of the tensor model is the combinatorics of the different components of the Kronecker product. The high dimensionality and the distributed nature of the memory traces in associative memories assure the preservation of function in the event of damage or absence of both neurons and synapses. This early-known characteristic is greatly accentuated in the case of memories in which the inputs were previously processed by the Kronecker or tensor product [27]. No more than a small statistical sample of the components of the tensor product is needed for the model to function properly [46]. Note that this feature is shared by all hyperdimensional computing models and vector symbolic architectures with distributed coding (for an extension of this point with abundant bibliographic references, it can be consulted [47]). This sample of the potential

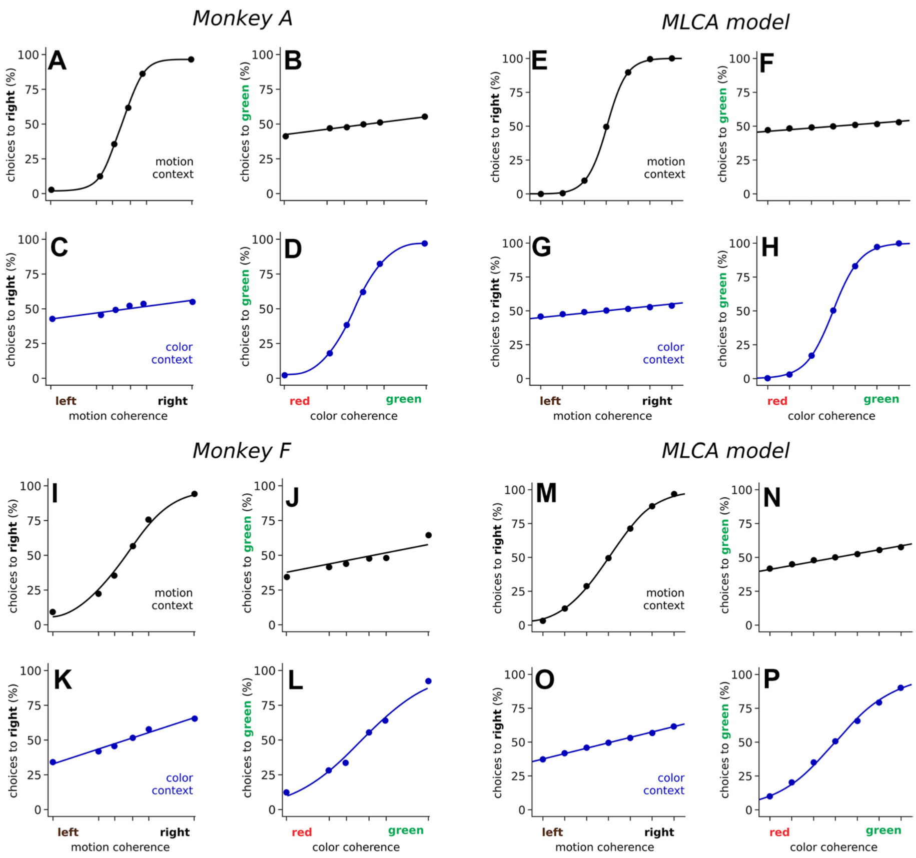

Fig. 4 Psychometric functions for monkeys A and F in [8] and for the MLCA model. Left panels (Monkey A and Monkey F) were adapted by permission from Springer Nature: Mante, V. et al. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84 (2013). For each, psychometric curves are obtained in the motion (top) and color (bottom) contexts as functions of motion (left) and color (right) coherences. Monkey A is globally better than monkey F in discriminating the task-relevant from the task-irrelevant attribute; the MLCA model is capable of capturing this difference by means of stimuli representation vectors that can be more linearly independent (monkey A) or less (monkey F). Our model shows a psychometric dependence on the task-irrelevant attribute qualitatively equivalent to that of the monkeys, an effect not shown by the recurrent neural network model of Mante et al.

combinatorics of the meeting of two neural afferences can be carried out on the dendritic tree of a neuronal group that carries memory. In addition, it is important to note that the processing of multiple Kronecker products, as those that were presented in the more general equations of the model (5–7), are just an abstract mathematical representation of an operation that can be carried out successively, by means of a neural recursion.

The other type of neural requirement is the kind of interaction between the afferent axons on a dendrite, which in the model is represented by a multiplication. Recent work has provided evidence for cortical functional flexibility at multiple scales [48], the type of operations that context-dependent memory models enable. The evidence points to inhibitory neurons underlying online functional sculpting of cortical circuits. At the level of an individual neuron, evidence has been found of specific subtypes of dendrite-targeting inhibitory neurons that control individual modality pathways converging onto a neuron. Wang et al. [49] found that this pathway-specific gating can be performed by inhibitory neurons when inputs from different pathways cluster on a pyramidal neuron dendrite. Their results suggest that context targets specific classes of inhibitory neurons. Notice that, mathematically, an inhibition or a functional shutdown can be modeled by a multiplication by zero.

This context-dependent gating-on mechanisms can also be found at the large scale of cortico-cortical networks, probably mediated by heterogeneous patterns of cholinergic signaling across the cortex [50]. Evidence points also to a key role of GABAergic interneuron populations in mediating mode-switching in cortical networks [49]. This type of scaffolding of connectivity across cortex areas can be modeled by input–output context-dependent tensor memories [39].

Hence, tensor-contexts matrix memories seem to be modeling exactly what the most recent experimental findings are revealing about functional flexibility at different levels of organization. Interestingly, tensor product nets became recently part of the computational weapons of deep learning models (see, for example, [51] and [52]). Eventually this could become another example of natural and cultural convergent evolution of computational solutions.

Finally, we want to point out that the natural continuation of this work is the representation of other situations where the information received contains elements that interfere with the correct answer for the proposed task. In particular, we are currently working on situations where the stimulus location interferes with the spatial, lateralized representation of the response, as it happens in the Simon and SNARC effects. In a series of experiments from the late 1960s, Simon and colleagues study a stimulus–response compatibility effect where subjects have to press one of two lateralized buttons in response to a visual or auditory stimulus, such as a color or a tone [53]. The stimulus is also lateralized. The Simon effect consists of larger response times and error rates when the stimulus and response are on opposite sides. The SNARC effect [54] consists of faster left-side responses to small numbers and faster right-side responses to large numbers, suggesting a spatial organization of magnitude information.

Appendix 1 Kronecker product

The Kronecker product, $\otimes$ (see [32]), is a particular case of the tensor product of two matrices of arbitrary dimensions, where $\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}$

The following are some of its basic properties:

(a) $\lambda(\boldsymbol{A} \otimes \boldsymbol{B}) = \boldsymbol{A} \otimes (\lambda \boldsymbol{B})$
(b) $\boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C})$
(c) $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{AC}) \otimes (\boldsymbol{BD})$
(d) $(\boldsymbol{A} \otimes \boldsymbol{B})^T = (\boldsymbol{A}^T \otimes \boldsymbol{B}^T)$

Of importance for context-dependent memory models, if **a**, **b**, **c**, and **d** are k-dimensional column vectors, using properties (c) and (d), the following equations hold:

$$(\mathbf{a} \otimes \mathbf{b})^T (\mathbf{c} \otimes \mathbf{d}) = \left(\mathbf{a}^T \otimes \mathbf{c}\right)\left(\mathbf{b}^T \otimes \mathbf{d}\right) = \langle \mathbf{a}, \mathbf{c} \rangle \langle \mathbf{b}, \mathbf{d} \rangle$$

Appendix 2 Simulation parameters

The MLCA parameters used to simulate the monkeys' decisions during the experiment performed by Mante et al. [8], as described in Sect. 5, are the following:

leakage ($\alpha = 0.1$); lateral inhibition ($\beta = 0.1$); integration step ($dt = 0.1$); time normalization ($\tau = 1$); Gaussian noise variance ($\sigma^2 = 0.25$); threshold boundary ($b = 10$).

The numerical simulations are performed with the context-dependent associative memory $\mathbf{M}$ defined in (21). The response and cue vectors are defined as orthogonal, while the motion and color vectors and the learning rates are fitted to the data; all vectors are normalized. The context-dependent associative memory for monkey A of [8] is defined by the vectors:

$\mathbf{r}_{left} = [1, 0]^T$; $\mathbf{r}_{right} = [0, 1]^T$; $\mathbf{c}_{motion} = [1, 0]^T$; $\mathbf{c}_{color} = [0, 1]^T$;
$\mathbf{p}_{left} = [0.9428,\ 0.3333]^T$; $\mathbf{p}_{right} = [0.3333, 0.9428]^T$;
$\mathbf{q}_{green} = [0.9129,\ 0.4082]^T$; $\mathbf{p}_{red} = [0.4082, 0.9129]^T$;

$$\eta = 1\ ; \quad \eta' = 0.8$$

and the context-dependent associative memory for monkey F is defined by the same vectors except for the context cues, which include a small overlap to capture this monkey's worse results:

$$\mathbf{c}_{motion} = [1, 0]^T; \quad \mathbf{c}_{color} = [0.25,\ 0.9682]^T$$

The numerical simulations following the dynamics of (20) are performed with the same context cues used for the memories, while the stimuli are varied analogously to Fig. 2B across 11 coherence levels:

$$\mathbf{p}_{motion} = \left[\sqrt{1 - i/10}, \sqrt{i/10}\right]^T; \quad i \in \{0, 1, \dots, 10\}$$

$$\mathbf{q}_{color} = \left[\sqrt{1 - i/10}, \sqrt{i/10}\right]^T; \quad i \in \{0, 1, \dots, 10\}$$

Acknowledgements We acknowledge the financial support by the Universidad de la República, Uruguay, and a postgraduate fellowship from the National Agency for Investigation and Innovation (ANII), Uruguay, to FML.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Marder, E.: Modifiability of pattern generation. Curr. Opin. Neurobiol. **1**, 571–576 (1991)
2. Chettih, S.N., Harvey, C.D.: Single-neuron perturbations reveal feature-specific competition in V1. Nature **567**, 334–340 (2019)
3. Clancy, K.B., Orsolic, I., Mrsic-Flogel, T.D.: Locomotion-dependent remapping of distributed cortical networks. Nat. Neurosci. **22**, 778–786 (2019)
4. Abbott, L.F.: Where are the switches on this thing? In: van Hemmen, J.L., Sejnowski, T.J. (eds.) 23 Problems in Systems Neuroscience, pp. 423–431. Oxford University Press, New York (2006)

5. Cichocki, A., Amari, S.I.: Adaptive blind signal and image processing. John Wiley & Sons (2002)
6. Braddick, O.: A short-range process in apparent motion. Vision Res. **14**, 519–527 (1974)
7. Baker, C.L., Jr., Braddick, O.J.: The basis of area and dot number effects in random dot motion perception. Vision Res. **22**, 1253–1259 (1982)
8. Mante, V., Sussillo, D., Shenoy, K.V., Newsome, W.T.: Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature **503**, 78–84 (2013)
9. Mizraji, E.: Context-dependent associations in linear distributed memories. Bull. Math. Biol. **51**, 195–205 (1989)
10. Mizraji, E.: Neural memories and search engines. Int. J. Gen. Syst. **37**, 715–738 (2008)
11. Smolensky, P.: Tensor product variable binding and the representation of symbolic structures in connectionist systems. Artif. Intell. **46**, 159–216 (1990)
12. Mizraji, E., Pomi, A., Valle-Lisboa, J.C.: Dynamic searching in the brain. Cognitive Neurodynamics **3**, 401–414 (2009)
13. Mizraji, E., Lin, J.: Logic in a dynamic brain. Bull. Math. Biol. **73**, 373–397 (2011)
14. Plate, T.A.: Distributed representations and nested compositional structure. University of Toronto, Department of Computer Science (1994)
15. Gayler, R.W.: Multiplicative binding, representation operators & analogy. In: Holyoak, K., Gentner, D., Kokinov, B. (eds) Advances in Analogy Research, p. 405. New Bulgarian University (1998)
16. Kanerva, P.: Hyperdimensional computing: an introduction to computing in distributed representation with high-dimensional random vectors. Cogn. Comput. **1**, 139–159 (2009)
17. Kohonen, T.: Associative Memory: a System-Theoretical approach Springer, New York (1977)
18. Longuet-Higgins, H.C.: Holographic model of temporal recall. Nature **217**, 104–104 (1968)
19. Gabor, D.: Holographic model of temporal recall. Nature **217**, 584 (1968)
20. Friston, K.J.: Functional and effective connectivity: a review. Brain Connect. **1**, 13–36 (2011)
21. Cooper, L.N.: Memories and memory: a physicist's approach to the brain. Int. J. Mod. Phys. A **15**, 4069–4082 (2000)
22. Cooper, L.N.: A possible organization of animal memory and learning. Proceedings of the Nobel Symposium on Collective Properties of Physical Systems. Aspensagarden (1973)
23. Anderson, J.A.: A simple neural network generating an interactive memory. Math. Biosci. **14**, 197–220 (1972)
24. Kohonen, T.: Correlation matrix memories. IEEE Trans. Comput. **C-21**, 353–359 (1972)
25. Anderson, J.A., Cooper, L.N., Nass, M., Freiberger, W., Grenander, U.: Some properties of a neural model for memory. AAAS Symposium Theoretical Biology and Biomathematics. AAAS, Washington, DC. (1972)
26. Anderson, J.A.: An Introduction to Neural Networks. The MIT Press, Cambridge (1995)
27. Mizraji, E., Pomi, A., Alvarez, F.: Multiplicative contexts in associative memories. BioSystems **32**, 145–161 (1994)
28. Anderson, J.A.: Associative networks. In: Arbib, M.A. (ed.) The Handbook of Brain Theory and Neural Networks, 2nd edn., pp. 117–122. MIT Press, Cambridge, MA (2003)
29. Minsky, M., Papert, S.: Perceptrons. MIT Press, Cambridge, MA (1969)
30. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986)
31. beim Graben, P., Potthast, R.: Inverse problems in dynamic cognitive modeling. Chaos Interdisciplin. J. Nonlinear Sci. **19**, 015103 (2009)
32. Smolensky, P., Legendre, G.: The Harmonic Mind: from Neural Computation to Optimality Theoretic Grammar. Volume 1: Cognitive Architecture. MIT Press, Cambridge, MA (2006)
33. Graham, A.: Kronecker Products and Matrix Calculus with Applications. Ellis Horwood, Chichester (1981)
34. Pomi, A., Mizraji, E.: Memories in context. BioSystems **50**, 173–188 (1999)
35. Mizraji, E., Lin, J.: Fuzzy decisions in modular neural networks. Int. J. Bifurcation Chaos **11**, 155–167 (2001)
36. Mizraji, E., Lin, J.: The dynamics of logical decisions: a neural network approach. Physica D **168**, 386–396 (2002)
37. Valle-Lisboa, J.C., Pomi, A., Cabana, A., Elvevåg, B., Mizraji, E.: A modular approach to language production: models and facts. Cortex **55**, 61–76 (2014)
38. Pomi, A.: Exploring the sources and mechanisms of cognitive errors in medical diagnosis with associative memory models. Diagnosis **4**, 251–259 (2017)
39. Pomi, A., Mizraji, E., Lin, J.: Tensor representation of topographically organized semantic spaces. Neural Comput. **30**, 3259–3280 (2018)

40. Usher, M., McClelland, J.L.: The time course of perceptual choice: the leaky, competing accumulator model. Psychol. Rev. **108**, 550–592 (2001)
41. Mazurek, M.E., Roitman, J.D., Ditterich, J., Shadlen, M.N.: A role for neural integrators in perceptual decision making. Cereb. Cortex **13**, 1257–1269 (2003)
42. Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., Brody, C.D.: Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature **520**, 220–223 (2015)
43. Bogacz, R., Usher, M., Zhang, J., McClelland, J.L.: Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. Philos. Trans. R. Soc. London, Ser. B **362**, 1655–1670 (2007)
44. Schall, J.D.: The neural selection and control of saccades by the frontal eye field. Philos. Trans. R. Soc. London, Ser. B **357**, 1073–1082 (2002)
45. Wang, X.J.: Probabilistic decision making by slow reverberation in cortical circuits. Neuron **36**, 955–968 (2002)
46. Valle-Lisboa, J.C., Reali, F., Anastasía, H., Mizraji, E.: Elman topology with sigma-pi units: an application to the modeling of verbal hallucinations in schizophrenia. Neural Netw. **18**, 863–877 (2005)
47. Gayler, R.W.: Vector symbolic architectures are a viable alternative for Jackendoff's challenges. Behav. Brain Sci. **29**, 78–79 (2006)
48. Cardin, J.A.: Functional flexibility in cortical circuits. Curr. Opin. Neurobiol. **58**, 175–180 (2019)
49. Yang, G.R., Murray, J.D., Wang, X.J.: A dendritic disinhibitory circuit mechanism for pathway-specific gating. Nat. Commun. **7**, 1–14 (2016)
50. Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A., Miller, K.D., Froemke, R.C.: Parallel processing by cortical inhibition enables context-dependent behavior. Nat. Neurosci. **20**, 62–71 (2017)
51. Cohen, N., Sharir, O., Shashua, A.: On the expressive power of deep learning: a tensor analysis. In: Conference on Learning Theory, pp. 698–728. PMLR. (2016)
52. Yu, D., Deng, L., Seide, F.: The deep tensor neural network with applications to large vocabulary speech recognition. IEEE Trans. Audio Speech Lang. Process. **21**, 388–396 (2012)
53. Simon, J.R., Small, A.M.: Processing auditory information: interference from an irrelevant cue. J. Appl. Psychol. **53**, 433 (1969)
54. Dehaene, S., Bossini, S., Giraux, P.: The mental representation of parity and number magnitude. J. Exp. Psychol. Gen. **122**, 371–396 (1993)