Blindfold learning of an accurate neural metric

Christophe Gardella[a,b], Olivier Marre[b,1,2], and Thierry Mora[a,1,2]

[a]Laboratoire de physique statistique, Centre National de la Recherche Scientifique, Sorbonne University, University Paris-Diderot, École normale supérieure, PSL University, 75005 Paris, France; and [b]Institut de la Vision, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Sorbonne University, 75012 Paris, France

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved February 23, 2018 (received for review October 26, 2017)

The brain has no direct access to physical stimuli but only to the spiking activity evoked in sensory organs. It is unclear how the brain can learn representations of the stimuli based on those noisy, correlated responses alone. Here we show how to build an accurate distance map of responses solely from the structure of the population activity of retinal ganglion cells. We introduce the Temporal Restricted Boltzmann Machine to learn the spatiotemporal structure of the population activity and use this model to define a distance between spike trains. We show that this metric outperforms existing neural distances at discriminating pairs of stimuli that are barely distinguishable. The proposed method provides a generic and biologically plausible way to learn to associate similar stimuli based on their spiking responses, without any other knowledge of these stimuli.

sensory discrimination | retina | neural metric | Restricted Boltzmann Machines | neural activity population models

A major challenge in neuroscience is to understand how the brain processes sensory stimuli. In particular, the brain must learn to group some stimuli in the same category and to discriminate others. Strikingly, this feat is achieved while the brain has only access to the noisy responses evoked in sensory organs but never to the stimulus itself. For example, the brain only receives the retinal responses to visual stimuli and is able to associate together responses corresponding to the same stimulus, while teasing apart the ones coming from distinguishable stimuli. How nervous systems can achieve such discrimination is still unclear. One strategy to solve this problem could be to learn either a decoding model to reconstruct the stimulus from the neural responses (1, 2) or an encoding model and invert it to find stimuli that can be distinguished (3). However, in both cases, this requires access to a lot of pairs of stimuli and evoked responses. Clearly, the brain is not guaranteed to have access to such data and may only access the neural response without knowing the corresponding stimulus.

Neural metrics (4), which define a distance between pairs of spike trains, have been proposed to solve this issue. In general, spike trains evoked by the same stimulus should be close by, while spike trains corresponding to very different stimuli should be far away. Using a given metric, one can associate together responses evoked by similar stimuli, without any information about the stimuli themselves (5, 6). The quality of this classification relies on the metric being well adapted to the task at hand, and different metrics are not expected to perform equally well.

Multiple metrics based on different features of the neural response have been proposed, mostly for single cells (7–12) and exceptionally for populations (13). These metrics do not use information about the correlative structure of the population response and often require tuning parameters to optimize performance, which requires external knowledge of the stimulus. In addition, a precise quantification of the performance of these different metrics at discriminating barely distinguishable stimuli is lacking.

Here we present an approach to learn a spike train metric with high discrimination capacity from the statistical structure of the population activity itself. We applied the method to the retina, a sensory system characterized by noisy, nonlinear (14), and correlated (15, 16) responses. We first introduce a statistical model of retinal responses, the Temporal Restricted Boltzmann Machine (TRBM), which allows us to learn an accurate description of spatiotemporal correlations in a population of 60 ganglion cells of the rat retina, stimulated by a randomly moving bar. We then use this model to derive a metric on neural responses. Using closed-loop experiments, where stimuli are tuned to be hardly distinguishable from each other, we show that this neural metric outperforms classical metrics at stimulus discrimination tasks. This high discrimination capacity is achieved despite the neural metric being trained with no information about the stimulus. We therefore suggest a general and biologically realistic method for the brain to learn to efficiently discriminate stimuli solely based on the output of sensory organs.

Results

Modeling Synchronous Population Activity with RBMs. We analyzed previously published ex vivo recordings from rat retinal ganglion cells (17). A population of 60 cells was stimulated with a moving bar and recorded with a multielectrode array (Fig. 1). Responses were binarized in 20 ms time bins, with value $\sigma_{it} = 1$ if neuron i spiked during a given time bin t, and 0 otherwise (Fig. 1). We first aimed to describe the collective statistics of spikes and silences in the retinal population, with no regard for the sequence of stimuli that evoked them.

We modeled synchronous correlations between neurons using RBMs (18, 19), which have previously been applied to retinal (20)* and cortical (21) populations. They give the probability of same-time spikewords $(\sigma_i) = (\sigma_{it})_i$ at any t as:

$$P[(\sigma_i)] = \frac{1}{Z}\sum_{(h_j)} \exp\left(\sum_i a_i\sigma_i + \sum_j b_j h_j + \sum_{i,j} W_{ji}\sigma_i h_j\right) \quad [1]$$

Significance

To understand how neural signals code sensory stimuli, most approaches require knowing both the true stimulus and the neural response. The brain, however, only has access to the neural signals put out by sensory organs. How can it learn to relate neural responses to sensory stimuli, especially for responses to which it has never been exposed? Here we show how to solve this problem by building a metric on neural responses such that responses to the same stimulus are close. Although the metric is built with no access to the stimulus, it outperforms all existing metrics in fine discrimination tasks, suggesting a way the brain could make sense of its sensory output.

Author contributions: C.G., O.M., and T.M. designed research; C.G. performed research; C.G., O.M., and T.M. contributed new reagents/analytic tools; C.G., O.M., and T.M. analyzed data; and C.G., O.M., and T.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]O.M. and T.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: tmora@lps.ens.fr or olivier.marre@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1718710115/-/DCSupplemental.

Published online March 12, 2018.

*Schwab DJ, Simmons KD, Prentice JS, Balasubramanian V, Computational and Systems Neuroscience (Cosyne) 2013, February 28-March 3, 2013, Salt Lake City.

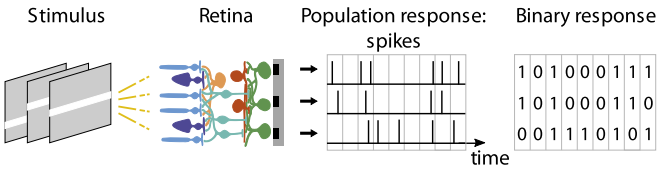

Fig. 1. Experimental setup. A rat retina is stimulated with a moving bar. Retinal ganglion cells (in green) are recorded with a multielectrode array. To model the response, spike trains are binarized in 20 ms time bins.

RBMs do not have direct interactions between neurons. Rather, their correlations are explained by interactions with binary latent variables, h_j, called hidden units (Fig. 2*A*). When a hidden unit takes a value of 1, it induces collective changes in the excitability of subpopulations of cells. Although it is tempting to think of hidden units as nonvisible neurons, they are only effective variables and usually do not correspond to actual neurons; hidden units can in fact reflect multiple causes of correlations, such as direct input from neighboring cells, or common input from intermediate layers and the stimulus. Their number can be varied: The more hidden units, the more complex structures can be reproduced by the model, but the more parameters need to be estimated.

We learned an RBM with 20 hidden units to model the responses of the retinal population responding to a randomly moving bar. The model was inferred on a training set (80% of responses) using persistent contrastive divergence (*SI Text*) and its predictions compared with a testing set (20% remaining responses). The RBM predicted well each neuron's firing rate (Fig. 2*B*) as well as correlations between pairs of neurons (Fig. 2*D*). In addition, the RBM predicted higher order correlations accurately, such as the distribution of the total number of spikes in the population (Fig. 2*C*). By contrast, a model of independent neurons (zero hidden units) underestimated the probability of events with few or many spikes by an order of magnitude. The model performance, measured by either the fraction of variance explained of pairwise correlations (Fig. 2*E*) or by the model log-likelihood (Fig. 2*F*), quickly saturated with the number of hidden units, with 15 units already providing near optimal performance.

TRBMs for Population Spike Trains. The RBM performs well at modeling neural responses within 20 ms time bins, but correlations between neurons often span longer time scales. To evaluate the importance of these longer term correlations, we plotted the distribution of the number of spikes in the population in a 100 ms time window (using the testing set) and compared it to the prediction from the RBM, where the response of the population was generated in each of the five 20-ms bins independently (Fig. 3*C*). Although the RBM performed better than a model of independent neurons, it still underestimated the probability of large numbers of spikes by an order of magnitude, indicating that correlations over longer scales than 20 ms play an important role in shaping the collective response statistics.

To account for these temporal correlations, we introduced the TRBM. This model generalizes the RBM by allowing for interactions between neurons and hidden units across different time bins (Fig. 3*A* and *SI Text*):

$$P[(\sigma_{it})] = \frac{1}{Z}\sum_{(h_{jt'})} e^{\sum_{it} a_i \sigma_{it} + \sum_{jt'} b_j h_{jt'} + \sum_{ijtt'} W_{ji,t'-t}\sigma_{it} h_{jt'}} \quad [2]$$

Because we want to describe the stationary distribution of spike trains regardless of the stimulus, absolute time is irrelevant, and the model is invariant to time translations: Connections between a hidden unit and a neuron only depend on the relative delay $t' - t$ between them. This property is similar to convolutional networks used in image processing but here in time instead of space.

We trained a TRBM with 10 hidden units per time bin, each connected to neurons across five consecutive time bins, on the same training set as before using persistent contrastive divergence (*SI Text*; see Fig. S4 for the inferred W couplings between hidden units and neurons) and compared predictions to the testing set. Like the RBM, the TRBM could predict individual neuron firing rates (Fig. 3*B*) and synchronous pairwise correlations (Fig. 3*D*). In addition, the TRBM could also predict temporal correlations ignored by the RBM. In particular, it reproduced accurately the distribution of the total number of spikes in a 100 ms time window, which the RBM did not (Fig. 3*C*). We also tested if the TRBM could predict correlations between the spiking activity of pairs of neurons in two time bins separated by a given delay. To do so, we computed the total variance of pairwise correlations for each delay and estimated the fraction of it that could be explained by the TRBM (Fig. 3*E* and *SI Text*). Even though direct connections between neurons and hidden units were limited to 80 ms, the TRBM could explain a substantial amount of correlations even for large delays, up to 150 ms where correlations vanish.

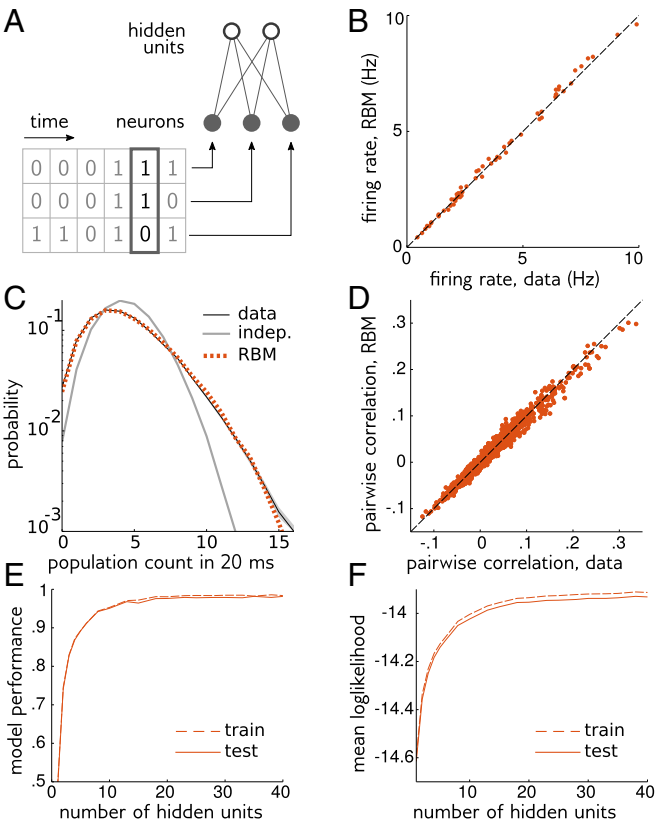

Fig. 2. The RBM model predicts accurately response statistics within single 20 ms time bins. (*A*) The RBM models the probability of binarized responses in single time bins. There are no direct interactions between neurons (gray circles). Instead, neurons interact with hidden units (white circles). (*B*) Single cell firing rate. Each dot represents the spiking frequency of a neuron in the testing set (not used for learning) versus RBM model prediction. (*C*) Distribution of the total number of spikes in the population during a time bin in the testing set (black) versus the prediction of a model of independent neurons (gray) or by the RBM (dotted red). Shaded area shows SE in data. (*D*) Pairwise correlations. Each dot represents the Pearson correlation for a pair of neurons in the testing set versus RBM prediction. (*E*) Fraction of the variance of correlations explained by RBM models, for different numbers of hidden units, in the training and testing sets. (*F*) Mean model log-likelihood in-sample (dashed line) and out-of-sample (full line) as a function of the number of hidden units. The small difference between training and testing sets suggests that there is no overfitting. indep, independent.

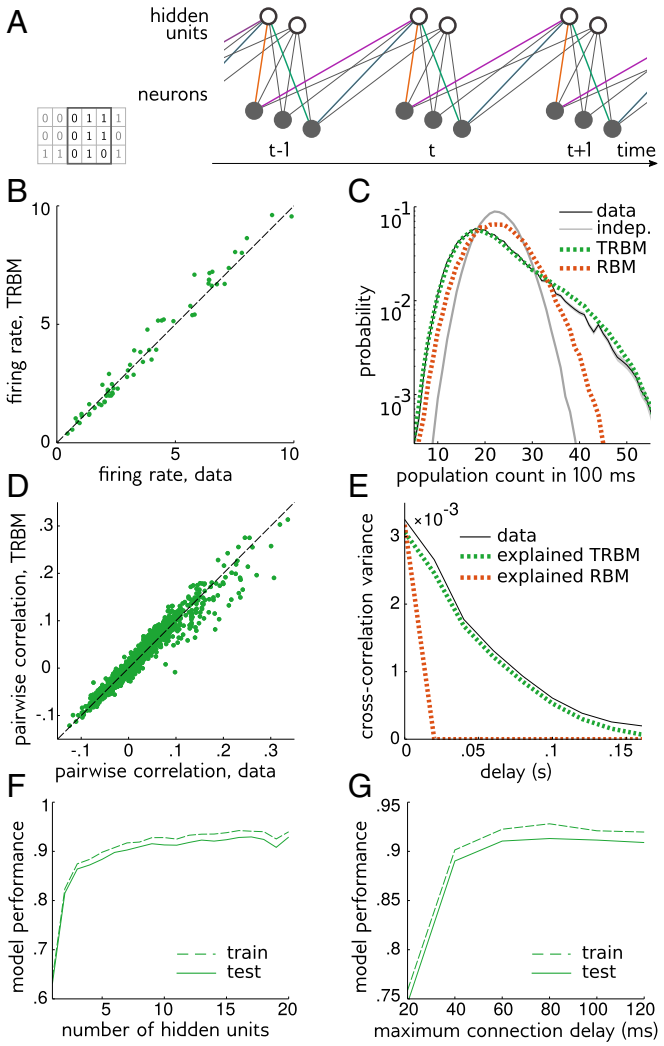

Fig. 3. The TRBM model predicts accurate response statistics across multiple time bins. (*A*) The TRBM's structure is similar to the RBM's, but neurons and hidden units are connected across multiple time bins. The interaction between neurons and hidden units only depends on the delay between them: In this schematic, interactions with the same color are equal. For simplicity, the model represented here only has interactions for delay 0 and 1 time bins. In general, there can be interactions with larger time delays. (*B*) Single cell firing rates; same as Fig. 2*B* but for the TRBM model. (*C*) Distribution of the number of spikes in the population during a 100 ms time window (five consecutive time bins), in the testing set (black), predicted by a model of independent time bins and independent neurons (gray), a model with independent RBMs in each time bin (dotted red), or a TRBM (dotted green). Shaded area shows SE in data. (*D*) Pairwise correlation. Same as Fig. 2*D* but for the TRBM model. (*E*) Cross-correlation. Black line shows the variance in cross-correlations between neurons with different time delays. Red and green lines show variance explained by RBM and TRBM, respectively. (*F*) Fraction of the variance of cross-correlations between neurons with delays up to 140 ms explained by TRBM models, as a function of the number of hidden units, in the training and the testing sets. (*G*) Same as *F* but varying the maximum connection delay between hidden and visible units. indep, independent.

Similarly to the RBM, we found that increasing the number of hidden units only marginally improved performance (as measured by the fraction of explained variance of pairwise correlations) beyond 10 units per time bin (Fig. 3*F*). We also varied the maximum connection delay between neurons and hidden units from 20 ms to 120 ms. Performance quickly saturated at a connection delay of around 60 ms (Fig. 3*G*). In the following, we will consider a TRBM with 10 hidden units and connection delay of 80 ms, unless mentioned otherwise.

A Neural Metric Based on Response Statistics. The hidden units of the TRBM can be considered as a way to compress the variability present in the neural activity and extract its most relevant dimensions. We asked whether these hidden units could be used to define a neural metric that would follow the structure of the population code, allowing for efficient discrimination and classification properties.

To this end, we designed neural metrics derived from the RBM and TRBM based on the difference between the hidden unit states. Take two responses, $\boldsymbol{\sigma} = (\sigma_i)$ and $\boldsymbol{\sigma}' = (\sigma'_i)$, of the retina and define $\boldsymbol{\Delta h} = (\Delta h_j)$ as the difference of mean value of the hidden units conditioned on the two responses, $\Delta h_j = \langle h_j \rangle_{\boldsymbol{\sigma}} - \langle h_j \rangle_{\boldsymbol{\sigma}'}$ (*SI Text*). Then the RBM metric is defined as:

$$d_{\mathrm{RBM}} = \boldsymbol{\Delta h}^{\mathsf{T}} \boldsymbol{W C W}^{\mathsf{T}} \boldsymbol{\Delta h}, \qquad [3]$$

where $\boldsymbol{C} = \langle \boldsymbol{\sigma}\boldsymbol{\sigma}^{\mathsf{T}} \rangle - \langle \boldsymbol{\sigma} \rangle \langle \boldsymbol{\sigma}^{\mathsf{T}} \rangle$ is the covariance matrix of the response and $\boldsymbol{W} = (W_{ji})$ is the matrix of couplings between neurons and hidden units. This definition can readily be generalized to the TRBM by adding time indices (*SI Text*). Note that this metric differs from the standard Euclidian distance in the space of hidden units, $\boldsymbol{\Delta h}^{\mathsf{T}} \boldsymbol{\Delta h}$: It has a nontrivial kernel $\boldsymbol{WCW}^{\mathsf{T}}$, which modulates the contribution of each hidden unit by its impact on neural activity. We will see later that this kernel improves discrimination capabilities. Note that this metric was defined without any information about the stimulus and solely from the knowledge of the activity. We next aimed to test how well this metric can discriminate pairs of stimuli.

Distinguishing Close Stimuli. To evaluate the capacity of a neural metric to finely resolve stimuli based on the sensory response, we introduce a measure of discriminability between the responses to two distinct stimuli based on neural metrics.

The response to a given stimulus is intrinsically noisy. Two repetitions of the same stimulus (let us call it reference stimulus) will give rise to two distinct responses, R_{ref} and R'_{ref}. The response R_{pert} to a perturbation of the reference stimulus may thus be hard to tease apart from another response to the reference stimulus, because of this noise (Fig. 4*A*). Given a neural metric $d(R, R')$, it is natural to define the discriminability of a perturbation as the probability for the response R_{pert} to be further apart from a response to the reference, R''_{ref}, than would two responses to the reference, R_{ref} and R'_{ref}, from each other:

$$\mathrm{Discr} = P(d(R''_{\mathrm{ref}}, R_{\mathrm{pert}}) > d(R_{\mathrm{ref}}, R'_{\mathrm{ref}})). \qquad [4]$$

If a perturbation is perfectly discriminable (Fig. 4*A*, *Left*), distances between reference and perturbation are well separated from distances within responses to the reference, and the discriminability will approach 1. Conversely, for perturbations too small to be discriminated, the two distributions greatly overlap (Fig. 4*A*, *Right*), and the discriminability is close to 0.5 corresponding to chance.

To finely assess the capacity of neural metrics to perform discrimination tasks, we need to study perturbations that lie between these two extremes, where discrimination is neither easy nor impossible. To find this soft spot, we performed closed-loop experiments where at each step the discriminability of a perturbation was analyzed to generate the perturbation at the next step (Fig. 4*B*; see ref. 17 for more details). We first recorded multiple responses to a reference stimulus, a 0.9 s snippet of bar trajectory described earlier (Fig. S1 *A* and *B*). We then recorded responses to many perturbations of this stimulus (Fig. 4*C*). For a given "shape" of the perturbation (i.e., normalized difference of bar position between reference and perturbation as a function of time; Fig. S1*C*), we adapted the perturbation size online and searched for the smallest perturbations that were still discriminable (*SI Text*). If a perturbation had high discriminability (as defined by a linear

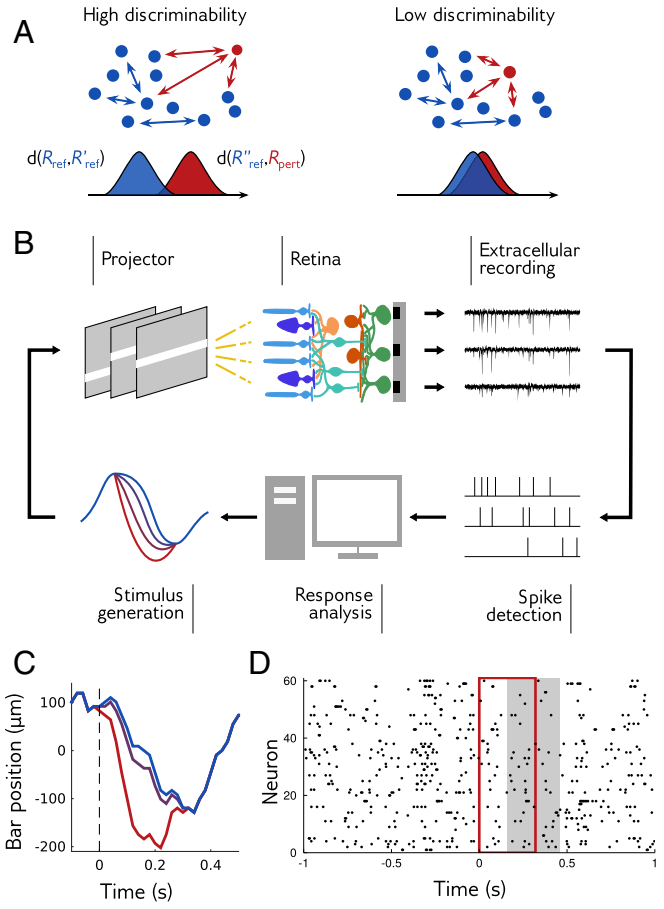

Fig. 4. Online adaptation of perturbations. (*A*) Discriminating with metrics. Stimulus discrimination is evaluated by comparing the distance of responses within the same reference stimulus (blue dots) and between the reference and a perturbation (red dots). Discrimability is defined as the probability that a within-stimulus distance (blue distribution) is lower than an across-stimuli distance (red distribution). (*B*) Closed-loop experiment. At each step, the rat retina was stimulated with a perturbation of a reference stimulus. Retinal ganglion cell responses were recorded extracellularly with a multi-electrode array. Electrode signals were high pass-filtered, and spikes were detected by threshold crossing. We computed the discriminability of the population response and adapted the amplitude of the next perturbation. (*C*) The stimulus consisted of repetitions of a reference stimulus (here the trajectory of a bar, in blue) and of perturbations of this reference stimulus of different shapes and amplitudes (see Fig. S1). Purple and red trajectories are perturbations with the same shape, at small and large amplitude. (*D*) Example population response. Each spike is represented by a dot. Red rectangle, duration of the perturbation; shaded rectangle, duration of responses for which the discriminability was measured.

discrimination task on the thresholded values of the raw multi-electrode array output, independently of any metric; see *SI Text*), at the next step we tested a perturbation with a smaller amplitude. Conversely, if a perturbation had low discriminability, we then tested a larger perturbation. Perturbations lasted 320 ms, and responses were analyzed over 300 ms with a delay (Fig. 4*D*).

Thanks to this method, we could explore the space of possible perturbations efficiently, exploring multiple directions (shapes) of the perturbation space simultaneously, and obtained a range of responses to pairs of stimuli that are challenging but not impossible to discriminate. This method allowed us to benchmark different metrics.

TRBM Metric Outperforms Other Neural Metrics at Fine Discrimination Tasks. We measured the discriminability (Eq. **4**) of a perturbation at different amplitudes, using the RBM and TRBM metrics (Fig. 5*A* and *SI Text*). As expected, the discriminability increased with the perturbation amplitude, with small perturbations being hardly discriminable from the reference stimulus (discriminability close to 0.5) and large perturbations almost perfectly discriminable (discriminability close to 1). Since this metric is based on the hidden states, it means that hidden states are informative about the stimulus. The much better performance of the TRBM, especially for small and medium perturbations, emphasizes the importance of temporal correlations in shaping the metric. For comparison, we computed the discriminability of the same perturbation for the Victor–Purpura metric (7) (*SI Text*), one of the first proposed neural metrics that has often been used in the literature to estimate the sensitivity of neural systems (22–24). This metric depends on a time scale parameter, which we optimized to maximize the mean discriminability of all recorded responses. Even with this optimization, the Victor–Purpura metric discriminated perturbations less well than either the RBM or TRBM metrics. Note that this optimization of the Victor–Purpura metric requires accessing pairs of stimulus responses, while the RBM and TRBM metrics never accessed this information. Despite this advantage given to the Victor-Purpura metric, it is still outperformed by the RBM and TRBM metrics.

To see if this better performance of our TRBM metrics held for other stimuli, we compared the discrimination capacity of the RBM and TRBM metrics with the Victor–Purpura metric for two different reference stimuli and 16 perturbation shapes for each (Fig. S1). For each reference stimulus and perturbation shape, we separated responses in batches of low, medium, and high discriminability, based on a linear discrimination task independent of any metric (*SI Text*, *E. Linear Discriminability*). We computed the mean discriminability of each response batch, for the RBM, TRBM, and Victor–Purpura metrics (Fig. 5 *B* and *C*). While responses in the low discriminability batch were poorly separated by all three metrics, a large majority of responses with medium and high discriminability had larger discriminability for the RBM metric, and even larger for the TRBM metric (Fig. 5 *B* and *C*), confirming the importance of temporal correlations.

We then compared the RBM and TRBM metrics to other neural metrics from the literature: van Rossum, angular, interspike interval (ISI), nearest neighbor, event synchronization, spike synchronization, and SPIKE metrics (definitions in *SI Text*), as well as the simple Hamming distance on the binarized responses. Metrics with free parameters were optimized to maximize their mean discriminability. Again, this optimization required accessing pairs of stimulus responses, while RBM and TRBM did not have access to this. For each metric, we computed the mean discriminability in each batch (low, medium, or high discriminability) across all reference stimuli and perturbation shapes (Fig. 5*C* and Fig. S3). Responses from the low-discriminability batch were hard to distinguish, and only five metrics did significantly better than chance ($p < 0.05$ for unpaired t test, Fig. S3): RBM, spike synchronization, SPIKE, and Angular and TRBM metrics. The TRBM metric discriminated responses the best and was significantly better than the second best, the SPIKE metric ($p = 0.014$, paired t test). For the medium- and high-discriminability batches, the RBM and TRBM metrics greatly outperformed all other metrics. Strikingly, in the medium-discriminability group, the improvement of discriminability above chance level was 30% higher for the RBM metric, and 94% for the TRBM metric, than for the Angular metric, the most discriminating metric from the literature.

This performance was little affected by the number of hidden units in the RBM and TRBM. The mean discriminability increases and eventually saturates with the number of hidden units (Fig. S2), indicating that the metric was not sensitive to that precise number, provided that it is large enough. By contrast, the TRBM-based metric using the unweighted Euclidian distance between the mean values of the hidden units degraded quickly with the number of units (dashed lines in Fig. S2). This worse performance may be explained by the fact that some

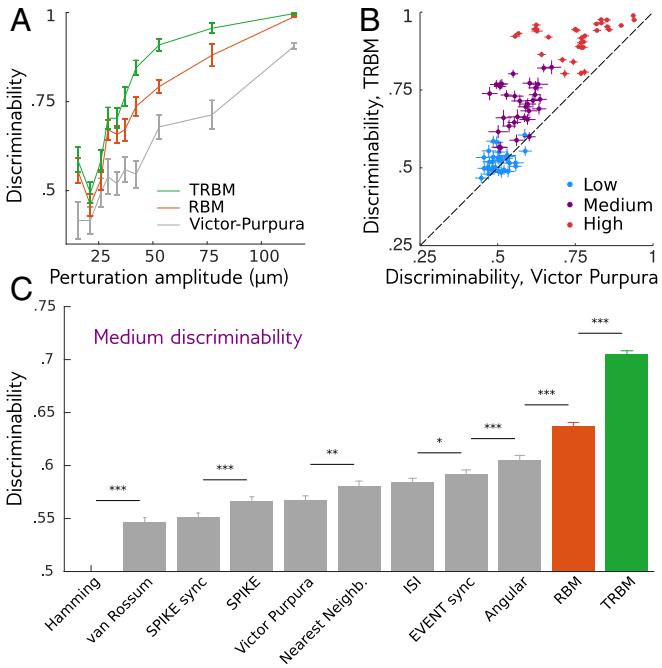

Fig. 5. RBM and TRBM metrics outperform classical metrics at discriminating responses. (*A*) Mean discriminability of responses to different amplitudes of an example perturbation shape, for the optimized Victor Purpura metric or the RBM and TRBM metrics. (*B*) Each point represents the mean discriminability for responses with low, medium, or high linear discriminability (*SI Text*), for one reference trajectory and one perturbation shape, for the Victor Purpura or TRBM metric. (*C*) Mean discriminability of responses with medium linear discriminability, across all reference stimuli and perturbation shapes. Stars on top of bars show significant difference in mean discriminability (paired *t* test, $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$). Error bars show standard errors in all three subpanels. Neighb, Neighbor.

hidden units have little or redundant impact on the activity but are counted with equal weight in the standard Euclidian distance. This stresses the importance of accounting for the impact of hidden units on the activity through a distance kernel as in Eq. **3**. We studied how the amount of data used for training the TRBM affected performance. We found that the TRBM metric reached near-saturating discriminability with as little as 40 s of activity and already outperformed other metrics with just 10 s of data (Fig. S6). We also checked that our conclusions were not affected by the choice of bin size (5, 10, or 40 ms).

Discussion

metric, the most discriminating metric fromWe have introduced a general method for building a metric from the neural responses of sensory organs, which outperforms all previously defined metrics at discriminating stimuli. Importantly, this TRBM metric is based on a statistical model of the activity that is trained in an unsupervised way, meaning that no knowledge of the stimulus is used. A previous approach used "semantic" metrics to cluster responses (3), but it required first learning a stimulus–response dictionary, which is impractical for the brain. By contrast, the TRBM metric emerges from the structure of the neural activity and does not use stimulus information, suggesting a realistic strategy for the brain to learn to discriminate stimuli. Many neural metrics require tuning parameters to maximize performance—a supervised task that uses stimulus information. Even after this optimization, the TRBM metric outperforms all metrics we found in the literature. This better performance can be explained by the TRBM's much larger number of parameters (3,070 in our example; see *SI Text*). However, this model complexity does not cause damaging overfitting, and near-optimal performance is achieved when training from just tens of seconds of activity (Fig. S6). Multineuron, many-parameter generalizations of existing metrics (13) could potentially achieve similar performance as the TRBM if properly trained, but it is not clear how to do it without supervision, especially on continuous stimulus ensembles where clustering (the archetypical unsupervised task) is not well-defined. Thus, the main advance of the TRBM metric over previous ones is its ability to be trained on any stimulus with no supervision. We checked explicitly that adding some degree of supervision to the TRBM did not improve discriminability: We optimized two directions of the TRBM parameter space (global rescalings of the couplings or of the hidden unit biases) and found little or no improvement (Fig. S7).

Although we have motivated the TRBM for defining a metric, statistical models of population activity deserve attention in their own right. In this regard, the TRBM provides an alternative to existing approaches that is both accurate and tractable. The spiking responses of retinal ganglion cells at a given time are strongly correlated (16, 25), and various strategies have been proposed to model their collective, synchronous (same time bin) activity. Central to this effort are Ising models (16, 26–29), also known as Boltzmann machines (30). These models are often hard to learn in practice and sometimes require additional terms (31, 32) or nonlinearities (33) to explain third- and higher order statistics such as the distribution of the total number of spikes. In Fig. S5 we show that an Ising model trained on our data indeed predicts higher order statistics worse than the RBM (see *SI Text*). It also is unclear how to exploit the structure of Ising models to derive a metric.

As an alternative to Ising models, RBMs were applied to the correlated activity in cortical microcolumns (21) and in the retina (20).† Our results confirm their ability to describe the synchronous collective activity in the retina, including pairwise correlations and the distribution of total numbers of spikes. Previous work also showed that the hidden units of a variant of the RBM conveyed information about the stimulus, although this was only made possible by the small number of used stimuli (34). All these models ignore correlations between spikes in different time bins, which play an important role as we have shown here. Maximum entropy models were generalized to account for correlations across time, but they were either practically intractable for large populations (35, 36) or only focused on the total number of spikes (37). The TRBM can reproduce pairwise and higher order correlations with high accuracy across different time bins, with a reasonable amount of parameters and relative computational ease. We therefore expect the TRBM to be useful in describing the stimulus-independent activity of neural populations in a variety of contexts (38–41).

Also using latent variables, a Hidden Markov Model was proposed to describe retinal activity (42), but it relied on a unique categorical variable, allowing for lower combinatorial diversity than the hidden units of our TRBM. Continuous latent variables have also been proposed to account for neural correlations in cortical networks (43), and it could be interesting to apply such models to the retina, although their training requires complex computational techniques. Our approach of using unsupervised learning (TRBM training) to inform a supervised task (stimulus discrimination) is reminiscent of the technique of unsupervised "pretraining" (44) used in machine learning when only a few examples are available. The link to machine learning suggests considering "deep" extensions of the RBM, with several layers of hidden variables (45), from which more general metrics could be derived. Deep (artificial) neural networks achieve higher discrimination power than RBMs when dealing with natural scenes and could lead to better metrics in our case as well.

The TRBM was trained on responses to the random motion of a bar and is specific to that stimulus ensemble. For instance, we

†Schwab DJ, Simmons KD, Prentice JS, Balasubramanian V, Computational and Systems Neuroscience (Cosyne) 2013, February 28-March 3, 2013, Salt Lake City.

checked that a TRBM metric trained on random checkerboards did worse than other metrics at discriminating bar trajectories. The brain may thus have to store several metrics for different stimulus ensembles or constantly relearn the metric depending on the visual stimulus. We have shown that this could be done within tens of seconds (Fig. S6), which is the typical adaptation time scale in the retina (46).

While neural metrics may not be explicitly estimated by the brain, our TRBM metrics have a natural biological implementation that suggests how a downstream population could discriminate responses to different stimuli. Hidden units could be implemented by a population of downstream neurons, with a simple response function: a weighted sum followed by a nonlinearity (see *SI Text*). This is reminiscent of a neuron summing responses from upstream cells, weighted by synapses' strengths, with delays to account for time lags. Indeed, it was shown that networks of spiking neurons can learn their synaptic weights to approximate RBMs (47). One can simplify our TRBM metric by linearizing the dependence of the hidden units as a function of activity (*SI Text*). Doing so leads to a metric that readily generalizes to continuous times, where the binning of time disappears. The metric is then simply given by a sum over pairs of spikes, with coefficients depending on the identity of the spiking neurons and the delay between them. We showed that this simplified "continuous" TRBM performs almost as well as the full TRBM metric (Fig. S3). The continuous TRBM metric could be implemented by simple summation of spikes with delays.

In summary, the TRBM provides insights into biologically possible representations of the stimulus with high discrimination capabilities, without the need for any supervised training. It would be interesting to compare the discrimination ability at the level of neural activity such as allowed by the TRBM with perceptual performance (48). However, since the relation between retinal activity and perception is indirect and affected by downstream processing (49), this issue is probably best tackled in cortical areas. None of the properties of the TRBM and its derived metric are expected to be specific to the retina, and our method could be readily applied to other sensory neural circuits.

Methods

Data sources and mathematical methods are presented in *SI Text*. Code is freely available at github.com/ChrisGll/RBM_TRBM.

Acknowledgments. We thank David Schwab for insightful discussions on RBMs and Ulisse Ferrari for making his code for learning Ising models available. This work was supported by Agence Nationale de la Recherche Grants "TRAJECTORY," "IRREVERSIBLE," and "Investissements d'Avenir" (ANR 10-LABX 65); European Commission Grant from the Human Brain Project (FP7-604102); Alliance pour les sciences de la vie et de la santé Grant "UNADEV"; and National Institutes of Health Grant U01NS090501.

1. Warland DK, Reinagel P, Meister M (1997) Decoding visual information from a population of retinal ganglion cells. *J Neurophysiol* 78:2336–2350.
2. Marre O, et al. (2015) High accuracy decoding of dynamical motion from a large retinal population. *PLoS Comput Biol* 11:e1004304.
3. Ganmor E, Segev R, Schneidman E (2015) A thesaurus for a neural population code. *eLife* 4:1–19.
4. Victor JD (2005) Spike train metrics. *Curr Opin Neurobiol* 15:585–592.
5. Machens CK, et al. (2003) Single auditory neurons rapidly discriminate conspecific communication signals. *Nat Neurosci* 6:341–342.
6. Narayan R, Graña G, Sen K (2006) Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol* 96:252–258.
7. Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: A metric-space analysis. *J Neurophysiol* 76:1310–1326.
8. Berry MJ, Warland DK, Meister M (1997) The structure and precision of retinal spike trains. *Proc Natl Acad Sci USA* 94:5411–5416.
9. van Rossum MC (2001) A novel spike distance. *Neural Comput* 13:751–763.
10. Quiroga RQ, Kreuz T, Grassberger P (2002) Event synchronization: A simple and fast method to measure synchronicity and time delay patterns. *Phys Rev E Stat Nonlinear Soft Matter Phys* 66:1–6.
11. Schreiber S, Fellous JM, Whitmer D, Tiesinga PHE, Sejnowski TJ (2003) A new correlation-based measure of spike timing reliability. *Neurocomputing* 52-54:925–931.
12. Hunter JD, Milton JG (2003) Amplitude and frequency dependence of spike timing: Implications for dynamic regulation. *J Neurophysiol* 90:387–394.
13. Houghton C, Sen K (2008) A new multineuron spike train metric. *Neural Comput* 20:1495–1511.
14. Gollisch T, Meister M (2010) Eye smarter than scientists believed: Neural computations in circuits of the retina. Nueron 65:150–164.
15. Arnett DW (1978) Statistical dependence between neighboring retinal ganglion cells in goldfish. *Exp Brain Res* 32:49–53.
16. Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
17. Ferrari U, Gardella C, Marre O, Mora T (2018) Closed-loop estimation of retinal network sensitivity reveals signature of efficient coding. eNeuro 4:e0166-17.2017.
18. Smolensky P (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds Rumelhart DE, McClelland JL, PDP Research Group C (MIT Press, Cambridge, MA, Vol 1, pp 194–281.
19. Hinton GE (2002) Training products of experts by minimizing contrastive divergence. *Neural Comput* 14:1771–1800.
20. Humplik J, Tkačik G (2016) Semiparametric energy-based probabilistic models. arXiv:1605.07371.
21. Köster U, Sohl-Dickstein J, Gray CM, Olshausen BA (2014) Modeling higher-order correlations within cortical Microcolumns. *PLoS Comput Biol* 10:1–12.
22. Aronov D, Reich DS, Mechler F, Victor JD (2003) Neural coding of spatial phase in v1 of the macaque monkey. *J Neurophysiol* 89:3304–3327.
23. Chase SM, Young ED (2006) Spike-timing codes enhance the representation of multiple simultaneous sound-localization cues in the inferior colliculus. *J Neurosci* 26:3889–3898.
24. Di Lorenzo PM, Chen JY, Victor JD (2009) Quality time: Representation of a multidimensional sensory domain through temporal coding. *J Neurosci* 29:9227–9238.
25. Schneidman E, Bialek W, Berry M, II (2003) Synergy, redundancy, and independence in population codes. *J Neurosci* 23:11539–11553.
26. Shlens J, et al. (2006) The structure of multi-neuron firing patterns in primate retina. *J Neurosci* 26:8254–8266.
27. Tang A, et al. (2008) A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J Neurosci* 28:505–518.
28. Shlens J, et al. (2009) The structure of large-scale synchronized firing in primate retina. *J Neurosci* 29:5022–5031.
29. Tavoni G, Ferrari U, Battaglia FP, Cocco S, Monasson R (2017) Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity. *Netw Neurosci* 1:275–301.
30. Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for Boltzmann machines. *Cogn Sci* 9:147–169.
31. Tkačik G, et al. (2014) Searching for collective behavior in a large network of sensory neurons. *PLoS Comput Biol* 10:e1003408.
32. Gardella C, Marre O, Mora T (2016) A tractable method for describing complex couplings between neurons and population rate. *eNeuro* 3:1–13.
33. Humplik J, Tkačik G (2017) Probabilistic models for neural populations that naturally capture global coupling and criticality. *PLOS Comput Biol* 13:e1005763.
34. Zanotto M, et al. (2017) Modeling retinal ganglion cell population activity with restricted Boltzmann machines. arXiv:1701.02898.
35. Vasquez JC, Marre O, Palacios aG, Berry MJ, Cessac B (2012) Gibbs distribution analysis of temporal correlations structure in retina ganglion cells. *J Physiol Paris* 106:120–127.
36. Nasser H, Marre O, Cessac B (2013) Spatio-temporal spike train analysis for large scale networks using the maximum entropy principle and Monte Carlo method. *J Stat Mech Theor Exp* 2013:P03006.
37. Mora T, Deny S, Marre O (2015) Dynamical criticality in the collective activity of a population of retinal neurons. *Phys Rev Lett* 114:1–5.
38. Yu S, Huang D, Singer W, Nikolić D (2008) A small world of neuronal synchrony. *Cereb Cortex* 18:2891–2901.
39. Kampa B (2011) Representation of visual scenes by local neuronal populations in layer 2/3 of mouse visual cortex. *Front Neural Circuits* 5:1–12.
40. Bathellier B, Ushakova L, Rumpel S (2012) Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* 76:435–449.
41. Truccolo W, et al. (2014) Neuronal ensemble synchrony during human focal seizures. *J Neurosci* 34:9927–9944.
42. Prentice JS, et al. (2016) Error-robust modes of the retinal population code. *PLoS Comput Biol* 12:e1005148.
43. Gao Y, Archer EW, Paninski L, Cunningham JP (2016) Linear dynamical neural population models through nonlinear embeddings. *Advances in Neural Information Processing Systems*, eds Lee DD, Sugiyama M, Luxburg UV, Guyon I, Garnett R (Curran Associates, Inc., Red Hook, NY), Vol 29, pp 163–171.
44. Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. *Neural Comput* 18:1527–1554.
45. Salakhutdinov R, Hinton G (2009) Deep Boltzmann machines. *Aistats* 1:448–455.
46. Wark B, Fairhall A, Rieke F (2009) Timescales of inference in visual adaptation. *Neuron* 61:750–761.
47. Nakano T, Otsuka M, Yoshimoto J, Doya K (2015) A spiking neural network model of model—Free reinforcement learning with high- dimensional sensory input and perceptual ambiguity. *PLoS One* 10:1–18.
48. Jacobs AL, et al. (2009) Ruling out and ruling in neural codes. *Proc Natl Acad Sci USA* 106:5936–5941.
49. Shapley R, Victor J (1986) Hyperacuity in cat retinal ganglion cells. *Science* 231:999–1002.