A universal power law optimizes energy and representation fidelity in visual adaptation

Matteo Mariani[1,✉], S. Amin Moosavi[1], Dario L. Ringach[1,2,*,✉] and Mario Dipoppa[1,*,✉]

[1]Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, USA
[2]Department of Psychology, University of California, Los Angeles, CA 90095, USA.
[*]Co-senior authors.
[✉]Corresponding authors: mattemari@ucla.edu; dario@ucla.edu; mdipoppa@ucla.edu

Abstract

Sensory systems continuously adapt their responses based on the probability of encountering a given stimulus. In the mouse primary visual cortex (V1), the average population response is a power law of the prior probability of stimuli in the environment. For a given stimulus type (e.g., oriented gratings), this power law is universal, with the same exponent observed across different statistical environments, enabling predictions of average population responses to new environments. Here, we aim to provide a normative explanation for the power law behavior. We develop an efficient coding model where neurons adjust their firing rates through multi-objective optimization, hypothesizing that the neural population adapts to enhance stimulus detection and discrimination while minimizing overall neural activity. We show that a power law that matches the one observed experimentally can emerge from our model. We interpret the exponent as reflecting a balance between energy efficiency and representational fidelity in adaptation. Furthermore, we account for the invariance of the power law's exponent across environmental changes by linking it to the dependence of tuning curve modulation on stimulus probability. Finally, we explain that variations in the exponent with different stimulus types (e.g., natural movies) result from changes in the minimal distances between neural representations, in agreement with experimental findings. We conclude that a universal power law of adaptation can be explained as a trade-off between representation fidelity and energy cost.

Introduction

Adapting to new environments is a hallmark of intelligent systems. Sensory systems continuously adapt their representation of stimuli based on environment statistics [1]–[7]. In the visual system, adaptation occurs across several areas of the brain, from the retina [8], [9] to the primary visual cortex (V1) [1], [2], [10]–[12] and in higher visual areas, such as V2 [11], V4 [13] and MT [14], involving different time scales, from milliseconds to seconds [6], [12], [15], [16] and different stimulus features, such as stimulus contrast [8], [10], [11], [16] and orientation [1], [17]–[19]. Adaptation also impacts our sensory experience, generating visual illusions in perception of orientations [20], [21] and higher visual features, such as faces [22].

Several theoretical frameworks, not always mutually exclusive, have been proposed to explain the mechanisms and functional benefits of adaptation [23]. These paradigms include redundancy reduction [24], [25], predictive coding [26]–[29], surprise salience [30], [31], Bayesian inference [32], [33], and efficient coding [34]–[38]. According to the efficient coding paradigm [39]–[42], neurons tend to maximize the amount of information about the environment they represent, subject to a limited amount of resources. Efficient coding has been particularly successful in explaining how neural populations adapt to the statistics of the environment [38].

Characterizing adaptation-induced changes in neural populations is crucial to fully understand how representations are transformed during adaptation [38], [43]. Experimental advances have enabled to study adaptation in populations of hundreds of neurons [1], [10], [38], [44]. A remarkable finding in one of these large-scale recording studies is that the average population response to a given stimulus is proportional to a power of its occurrence probability [1]. In this study, during a set time block, stimuli were sampled according to different probability distributions, which defined a statistical sensory environment for that time block. For a given stimulus type (e.g. gratings), the power law exponent is universal, i.e. approximately invariant across environments (Fig. 1a). However, the exponent depends on the stimulus type: its magnitude is smaller for natural movies than for oriented gratings (Fig. 1b).

In the present study, we introduce a normative model that reproduces this universal power law of adaptation using an efficient coding approach. We show that the power law arises as a trade-off between energy cost and representation fidelity, which we quantify as the ability to detect and discriminate visual stimuli. We also offer a geometric interpretation for the reduction of the exponent in natural movies, showing that it may be due to an increase in the minimal angular separation between neural response vectors evoked by the stimuli.

Results

We summarize the three main experimental results in [1] as follows. Let us first consider one environment e, defined, for a given stimulus type (e.g. an orientation θ), by the stimulus probability $p_e(\theta)$. We define, for a given stimulus θ and environment e, the response vector $\mathbf{r}_e(\theta) = \{r_{e,1}(\theta), r_{e,2}(\theta), \ldots, r_{e,n_R}(\theta)\}$, namely the vector of the firing rates of each of n_R neurons. Consider the average population response for the environment e, $r_e(\theta) = \frac{1}{n_R}\sum_{i=1}^{n_R} r_{e,i}(\theta)$. The first observation in [1] is that given two environments A and B (Fig. 1a), the ratio of the average population responses as a function of the average environment probabilities is described by the following power law:

$$\frac{r_A(\theta)}{r_B(\theta)} = \left[\frac{p_A(\theta)}{p_B(\theta)}\right]^{\beta}, \tag{1}$$

The second observation is that the exponent β is approximately invariant to the choices of A and B (Fig. 1a). The third observation is that β depends on the type of stimuli and is lower in magnitude for natural movies (Fig.1b) than for oriented gratings.

To explain these three observations, we introduced an efficient coding model. For the case of oriented gratings, let us consider a homogeneous set of reference tuning curves $r_0(\varphi, \theta)$, where φ indexes each neuron's preferred orientation and θ is the stimulus orientation, both parametrized between 0 and π. Tuning curves of different neurons differ only by translation, reflecting their distinct preferred orientations. Given an environment A, our model modifies $r_0(\varphi, \theta)$ to find the adapted tuning curves $r_A(\varphi, \theta)$ as the optimal solution of a scalarized multi-objective cost (Fig. 1c). Specifically, our model numerically finds a function $a(\varphi, \theta)$ modulating the reference tuning curves as follows: $r(\varphi, \theta) = r_0(\varphi, \theta)a(\varphi, \theta)$. The adapted tuning curves minimize the following total cost function:

$$\mathcal{C} = \lambda_E \mathcal{E} + \lambda_D \mathcal{D} + \lambda_T \mathcal{T} + \lambda_H \mathcal{H} + \lambda_S \mathcal{S}. \tag{2}$$

Each of these terms contributes to shaping the optimal tuning curves. The term $\mathcal{E}[p_A(\theta), r(\varphi, \theta)]$ is the expected *energy cost* (i.e. the metabolic consumption of firing action potentials), over the probability $p_A(\theta)$, of encoding the stimuli in the environment e. The energy cost penalizes a high average population response. The term $\mathcal{D}[p_A(\theta), r(\varphi, \theta)]$ is the expected *discrimination error* over all pairs of stimuli (θ_i, θ_j), weighted by $p_A(\theta_i) \cdot p_A(\theta_j)$. The discrimination error penalizes small distances between response vectors encoding different stimuli relative to the noise, which we modeled as independent, and identically distributed. The term $\mathcal{T}[p_A(\theta), r(\varphi, \theta)]$ represents the expected *detection error* across all stimuli, quantifying the ability to distinguish the presence of a stimulus from noise. This error small amplitudes of the average population responses relative to the noise. The *homeostasis constraint* $\mathcal{H}[r(\varphi, \theta)]$ induces the adapted tuning curves to be similar to the reference tuning curves $r_0(\varphi, \theta)$. The *smoothness constraint* $\mathcal{S}[r(\varphi, \theta)]$ enforces each neuron's tuning curve to be smooth as a function of the stimuli. The weights $\vec{\lambda} \equiv \{\lambda_\alpha : \lambda_\alpha > 0\}$, with $\alpha \in \{E, D, T, H, S\}$, are hyper-parameters of the model and represent the relative importance of each term in the optimization. A power law emerged for a wide range of these hyper-parameters (Supplementary Fig S.2). Thus, finding the adapted tuning curves to an environment A reduces to solving $a_A(\varphi, \theta) = \arg\min_{a(\varphi,\theta)} \mathcal{C}[\vec{\lambda}, p_A(\theta), r_0(\varphi, \theta), a(\varphi, \theta)]$. The optimization was performed iteratively by gradient descent subject to the following constraint: $a_A(\varphi, \theta) = a_{A,1}(\varphi)a_{A,2}(\theta)$.

The power law of adaptation emerges from an efficient coding model. Such an efficient coding model reproduced the structure of the tuning curves modulation. We first considered two environments: one defined by a uniform probability distribution and the other by a von Mises, or biased, probability distribution (Fig. 2a). We used the model to predict how the tuning curves adapt in both environments. As expected, the tuning curves in the uniform environment $r_U(\varphi, \theta)$ were homogeneous (Fig. 2b). Thus, they had the same structure of the reference tuning curves $r_0(\varphi, \theta)$, differing only in amplitude. We compared the tuning curves $r_U(\varphi, \theta)$ in the uniform environment (Fig. 2b) to the tuning curves $r_A(\varphi, \theta)$ in the biased environment (Fig. 2c). The ratio $r_A(\varphi, \theta)/r_U(\varphi, \theta)$ was modulated more strongly across stimuli than across neurons (Fig. 2d). This model's prediction was validated by a novel analysis of the experimental data in [1]: the ratio of tuning curves in a von Mises environment to those in a uniform environment, averaged across recording sessions, exhibited stronger modulation across stimuli than across neurons. (Fig. 2e).

A power law emerged as the solution of our efficient coding model. From the tuning curves, we next computed the average population responses for the uniform and biased environments $r_U(\theta)$ and $r_A(\theta)$ (Fig. 2f, inset). The ratio of the average population responses as a function of the ratio of probabilities followed almost exactly a power law (Fig. 2f) as in the experimental data [1]. Note that, because in our model the tuning curves are homogeneous in a uniform environment, we obtain a power law for the average population response as a function of the environment probability for any other environment e, namely $r_e(\theta) \propto p_e(\theta)^\beta$, without the need to consider the ratios of these quantities. However, in the experiments it is necessary to take such ratios to obtain a power law, because the tuning curves are inhomogeneous in a uniform environment. Our choice to consider homogeneous tuning curves in a uniform environment was motivated by our focus on the effect of short-term adaptation, abstracting away from inhomogeneities, which are likely the results of longer-term effects. We will later show that our efficient coding model reproduces the power law also in the case of inhomogeneous tuning curves (Figure 4).

The energy cost, detection error, and homeostasis constraint were necessary to obtain a power law. To see this, we set to zero, one at a time, each of the weights λ_α. For example, considering a von Mises environment (Supp. fig. S.1a), removing the energy cost led to a relation of the average population response which is inconsistent with the

experiments of [1] for two reasons: it deviates from a power law and is also an increasing function of the probability (Supp. fig. S.1b). Removing either the detection cost or the homeostasis constraint, did not lead to a the power law (Supp. fig. S.1c,d). Removing the homeostasis constraint flattens the adapted average population response to almost a constant (Supp. fig. S.1c). The presence of the homeostasis constraint maintains the average population response closer to the reference one, forcing the average population response to be a non-constant function of the probability. Removing the detection error results in an average population response deviating from a power law, decreasing more sharply as the probability of stimuli increases (Supp. fig. S.1d). A power law was preserved when any of the other two terms, smoothness and discrimination, was removed (Supp. fig. S.1e,f). To further understand the contribution of energy cost, detection error, and homeostasis constraint, we analyzed a simpler model including only these three terms. We confirmed that the power law is only approximate (as in Figure 2f), but there exists a region of the weights in which the approximation has a very low error (Supp. fig. S.2a-b-d) and the exponent changes smoothly from -1 to 1 (Supp. fig. S.2c-d-e), including values comparable with data. The power law is thus almost exact in a range of probabilities compatible with the experiments (Supp. fig. S.3). It is thus possible to obtain the power law for a wide range of each of the model's hyper-parameters, including the reference tuning curves' amplitude (Supplementary Fig. S.4).

The smoothness term was necessary to reproduce the continuity of average population responses in the discontinuous environments observed experimentally. Neural recordings in [1] show that even when the stimulus probability $p_e(\theta)$ changes abruptly, the average population responses $r_e(\theta)$ remain smooth (Supp. fig. S.1g). To capture this effect, enforcing smoothness in tuning curves is necessary; otherwise, discontinuities arise (Supp. fig. S.1l).

The discrimination term was necessary to capture the small but significant changes in response vector angles observed in the experiments. Specifically, in experiments [1], the response vectors $\mathbf{r}_A(\theta)$ and $\mathbf{r}_B(\theta)$ for two environments, A and B, were observed to be rotated relative to each other (Supp. Fig. S.1m). To ensure that the model reproduced the angular shift between adapted response vectors in different environments, the discrimination term could not be omitted (Supp. Fig. S.1q). Without the discrimination term, in fact, responses would scale uniformly across neurons, modifying only the magnitude of the response vector while preserving its direction. The discrimination term was also necessary to explain the change in the power law exponent for natural movies (Figure 5)

The power law exponent is independent of the environment and reflects the relative magnitude of the model's weights. The model predicted that the power law exponent is invariant to changes in the statistical environment. We studied the output of the model to several smooth environments (Fig. 3a) compatible with the experimental ones [1]. We sampled pairs of environments and observed that the relation between ratio of probabilities and ratio of average population responses was approximately invariant across pairs of environments, allowing us to fit all the points of the model with one power law (Fig. 3b). The exponent's invariance also emerged in the simpler model that only included the energy cost, detection error, and homeostasis constraint, introduced earlier: we analytically computed the modulation and found that it depends only on the stimulus probability p, not explicitly on the stimulus θ, (Supplementary Information, *The relation between average population responses and probabilities is invariant under*

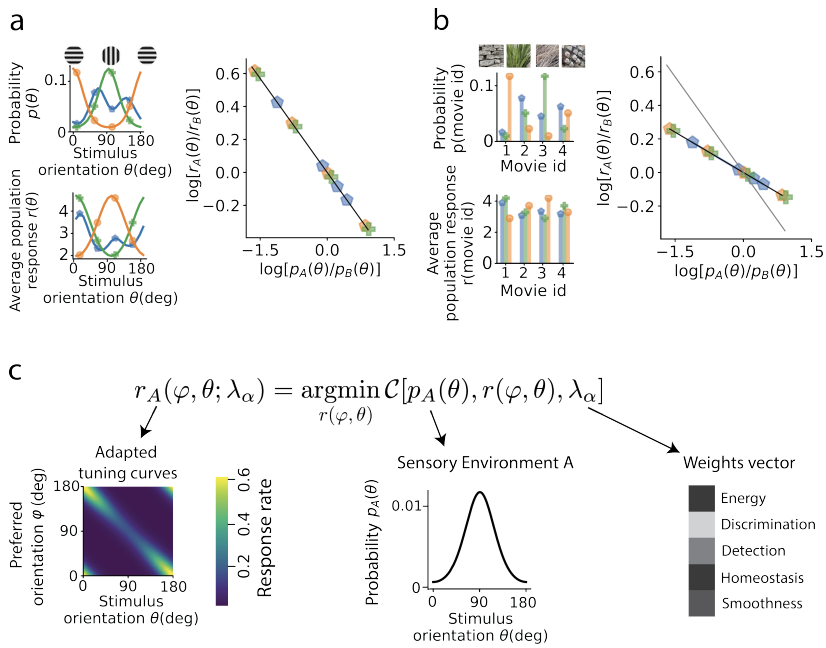

Figure 1: **An efficient coding model to explain a universal power law of visual adaption.** a) *Top left:* three environments of oriented gratings. Each environment is characterized by a probability distribution of the grating orientations. *Bottom left:* the average population responses to each of the environments in the top panel. The responses are paired with the environments by color. *Continues on the next page.*

Figure 1: *Continued caption: Right:* simplified representation of two key empirical findings from [1] that our model seeks to explain: (i) the ratio of average population responses between two different statistical environments, A and B, follows a power law of the ratio of the respective stimulus probabilities; (ii) the exponent of this power law remains invariant across different pairs of environments. Each colored marker represents a unique pair of environments; for clarity, only three pairs are shown, with colors indicating the environment in the numerator of the ratio. Each marker corresponds to one in the top left panel with same shape and color. Regardless of the environment pair, all points align along the same line, represented here in black. b) *Right:* simplified representation of a third key empirical finding in [1] that our model seeks to explain: the power law's exponent depends on the stimulus type, for example, it is smaller in magnitude for natural movies (colored markers and black line as in (a)) than for oriented gratings (reference gray line). *Top and bottom left:* same as (a), but for environments of natural movies. c) Efficient coding model: in a given environment, the tuning curves (bottom left, represented here with a heat map depicting the response rate as a function of stimulus θ and neuron's preferred orientation φ) are the solution of a multi-objective optimization of a total cost (top) depending on the environment $p_A(\theta)$ (bottom middle) and weights λ_α (bottom right). Each λ_α multiplies one of five different terms corresponding to the energy cost, discrimination error, detection error, homeostasis constraint, and smoothness constraint of the tuning curves.

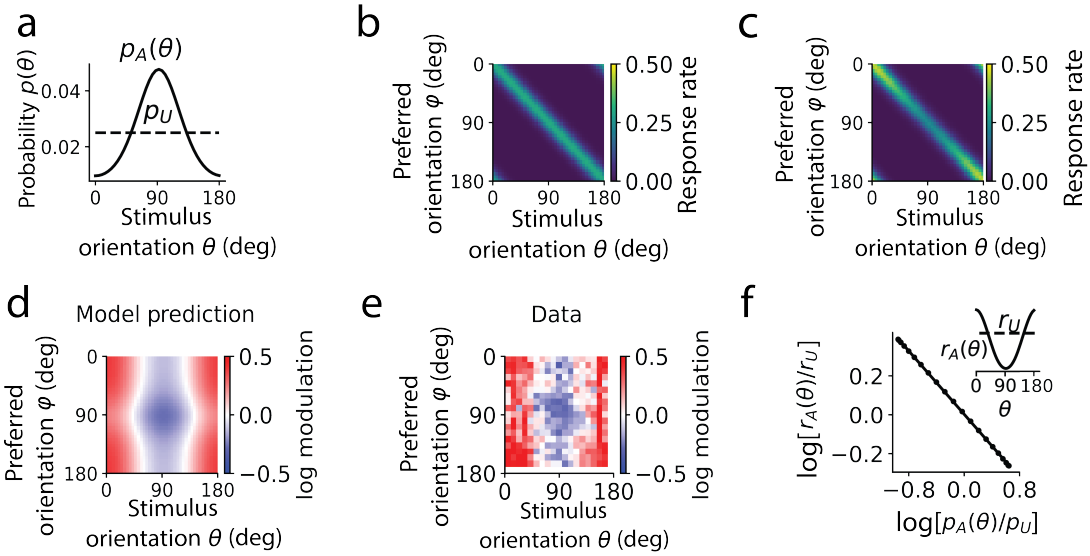

Figure 2: **The model predicts a power law of adaptation via a modulation of tuning curves consistent with the data.** a) Two example environments: a uniformly distributed environment (dashed line) and a von Mises distributed environment (solid line). b) Tuning curves, represented as a heat map depicting the response rate as a function of stimulus θ and neuron's preferred orientation φ, emerging from optimization of the total cost in Fig. 1c and in equation (2), in the uniform environment in (a). c) Similar to (b) but the environment is the von Mises in panel (a). d) Modulation of the tuning curves emerging from our model as the ratio between the tuning curves in (c) and in (b). e) Modulation in experimental data in [1] obtained by averaging groups of neurons with similar orientation preferences. f) Logarithm of the ratio of average population responses between the von Mises and uniform environments in (a) as a function of the logarithm of the ratio of their probabilities (circles) and the power law fit (solid line). *Inset:* the average population responses in the uniform (dashed line) and biased environment (solid line). In our model, rates are estimated using the L_1 norm of the rate vector, while [1] uses the L_2 norm; the two are however highly correlated in the data [1].

environment change). Thus, the adapted response rate is determined solely by the probability of a stimulus, regardless of the overall stimulus distribution. Consequently, there is only one (universal) function that relates the adapted rate to the probability and for a range of parameters this universal function turns out to be approximately a power law (Supplementary fig. S.2).

The power law exponent depended on the relative magnitude of the model's weights. For example, to characterize the dependence of the exponent on the energy cost, we focused on one example environment (Fig. 3c). Then, we varied the weight of the energy cost λ_E, keeping the other weights fixed. We did not fine-tune the other weights, since the power law was valid for a range of values of each weight λ_α. For sufficiently large values of the energy weight λ_E, the energy cost dominated the other costs in eq. (2), yielding a negative power law exponent. As a result, average population responses decreased with stimulus probability. This minimized overall neural activity by reserving high responses for rare stimuli. Conversely, for lower values of λ_E, the relative importance of the other terms increased, leading to a positive exponent (Fig. 3d). Consequently, the average population response increased with stimulus probability, enhancing detection and discrimination, and thus the representation fidelity, of the most probable stimuli.

The invariance of the power law exponent across environments was respected for a wide range of values of the energy weight. A variation in λ_E, changed the exponent of the power law of approximately the same amount for different environments (Fig. 3e). Thus, we confirmed the result observed in an example environment (in Fig. 3c-d): the change in the exponent is caused by a trade-off between conflicting constraints. Consistently with this result, detection and discrimination worsen when the energy spent is less (Fig. 3f).

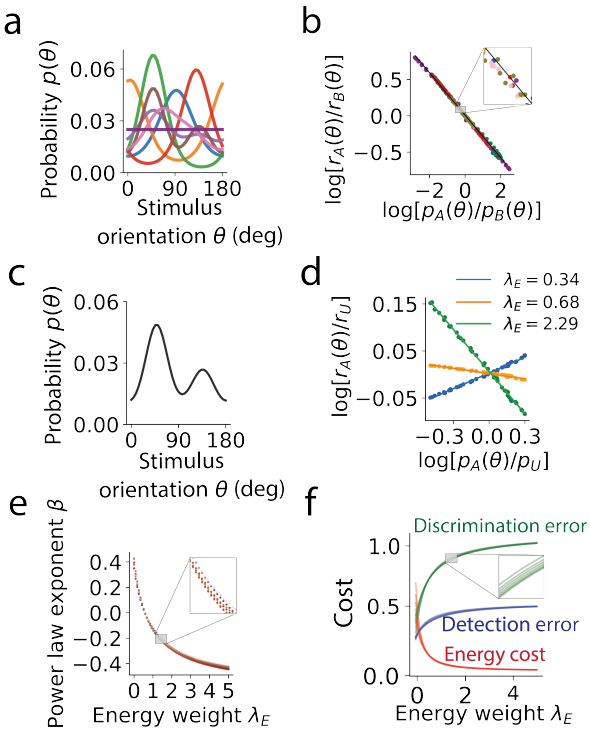

Figure 3: **The power law is universal and arises from a trade-off between representation fidelity and energy cost.** a) Set of example environments defined by different probability distributions of the stimulus orientation. b) Colored circles: ratios of response rates as a function of ratio of probabilities from different pairs of environments in (a); solid line: power law fit. Response rates were computed with a fixed weight vector λ_α. *Inset:* zoom around the origin to appreciate negligible variations across pairs of environments. c) One example environment from (a). d) Three outcomes of the model (dots) and their power law fits (solid lines) obtained as solutions of the total cost minimization for the environment in (c) at different levels of energy weight λ_E (all the other λ_α being fixed within broad ranges of values that guaranteed the power law). e) The power law exponent β as a function of the energy weight λ_E for each pair of environments in panel (a). Each color corresponds to a pair of environments in (a). *Inset:* zoom around exponent $\beta \approx -0.3$, which is the closest to experiments in [1], to appreciate negligible variations across pairs of environments. f) Energy cost, discrimination error, and detection error as functions of the energy weight λ_E for the different environments in (a). *Inset:* zoom of the discrimination error to appreciate negligible variations across environments.

The model explains the modulation of individual tuning curves in the data. Our model predicted the adaptation of the experimentally recorded tuning curves, which are not homogeneous in the uniform environment. Inhomogeneities in tuning curves lead to a bias in the average population response to some specific stimuli. The average population responses, therefore, are not power laws of the probability. In real data, it is the ratio of average population responses, which is a power law of the ratios of probabilities. This is because taking the ratios of the average population responses cancels out the biases present in the empirical data. To verify whether our model cancels out these biases, we modulated the individual tuning curves in [1]. Each recording session in [1] included three different environments (Figure 4a). As an example, we randomly chose one of the three environments and used the recorded tuning curves in that environment (Figure 4b) to train the model. Through gradient descent, we trained the model to find a set of weights $\vec{\lambda}^*$ ensuring that the adapted tuning curves, and consequently the average population responses, were as close as possible to those of the data (Figure 4c; Methods). Subsequently, we used these optimal weights to predict the average population responses in the two holdout test environments, finding a strong match between model predictions and experimental data (Figure 4d,e). This result suggests that the weight vector $\vec{\lambda}$ directly controls the modulation of tuning curves. For the homogeneous tuning curves in the model, fixing $\vec{\lambda}$ determines the power law exponent β (Figure 3e), which remains invariant across environments. Similarly, for heterogeneous neural responses, fixing $\vec{\lambda}$ dictates how probability modulates the tuning curves. We performed this analysis on all recording sessions in [1]. For each recording session, we iterated through environments as training and test sets. The exponents given by the model were correlated with the observed power law exponents in the real data (Supp. fig. S.5, Pearson Correlation Coefficient $\rho = 0.37$).

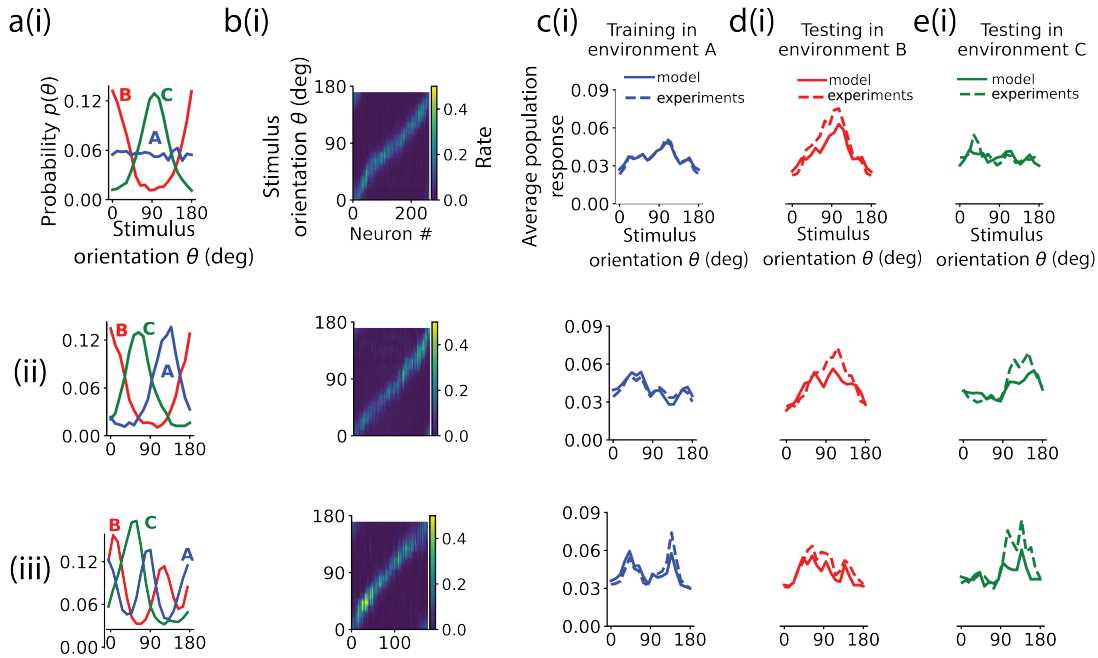

Figure 4: **The model fits individual tuning curves and predicts average population responses that align with experimental data.** a(i) three smooth environments of orientations from one example recording session in [1]. b(i) Experimental tuning curves of recorded neurons in the training environment A in panel a(i). c(i) average population response to stimuli in the experiments (dashed line) and the trained model (solid line). The model was trained to best fit the tuning curves shown in panel b(i). Note that in the model, as in the data, the uniform environment does not lead to a uniform average population response due to inhomogeneities in the tuning curves. d-e(i) Similar to c(i), but the average population response was for environments B and C in panel a(i). The model predictions on the two test environments B and C were obtained by using the weights λ_α from the training environment A. Rows (ii-iii): same as row (i), but with triples of environments progressively more complex from other two example recording sessions in [1].

Larger distances between neural representations explain shallow power law for natural movies. By analyzing the data in [1], we observed that the closest neural representations of natural movies were farther apart than the closest neural representations of oriented gratings. To understand why the exponent is closer to zero for natural movies (Fig. 1b), we analyzed angles between the high dimensional neural representations (i.e. the response vectors) for both natural movies and oriented gratings (Fig. 5a). First, we note that while oriented gratings can be described by a single parameter (orientation angle), natural movies are inherently high-dimensional. As a result, a finite set of randomly sampled natural movies is likely to be highly dissimilar, whereas a finite set of oriented gratings will inevitably include some that are relatively similar to each other. This difference between stimulus types leads to differences in response vectors. We computed the angles between any pair of response vectors $\alpha(x_i, x_j)$ from the data in [1]. Then, for each stimulus x_i, we selected only pairs such that the other stimulus x_j formed the smallest angle in the neural space (Figure 5b; see Methods). Comparing the two distributions we found that the minimal angle between response vectors, which we defined as $\alpha_{\min}$, was higher if the stimuli were natural movies, rather than oriented gratings. This minimal angle is greater than zero due to the finite number of stimuli.

A one-dimensional model of neural representations captured far apart neural representations (large $\alpha_{\min}$), the essential feature of natural movies described above. We introduced spaced environments, that we defined by a discrete probability distribution with a finite minimal distance $\Delta\theta_{\min}$ between stimuli (Fig. 5c). Our model can be used to study both continuous (Suppl. Fig. S.6a-c) and discrete (spaced) environments (Suppl. Fig. S.6b-c) of stimuli. In the spaced model, the angle α between response vectors depended not only on the angle $\Delta\theta$ between stimuli, but also on the sharpness κ_r of tuning curves (Fig.5d). More specifically, α saturated faster for large κ_r. Consequently, for a sufficiently large value of κ_r, an arbitrary value of the angle between response vectors α could be obtained for a given angle between stimuli $\Delta\theta$, since the two quantities are related by a function that gets steeper with κ_r.

The large minimum angles between neural representations of natural movies explained the small absolute value of the exponent for this type of stimuli. We computed the adapted responses to several environments differing by the minimal angle between stimuli $\Delta\theta_{\min}$. Increasing $\Delta\theta_{\min}$ led to an increase in the minimal angle between response vectors, $\alpha_{\min}$, which in turn caused the power-law exponent to change from $\beta = -1/3$ (as in oriented gratings) to $\beta = -1/5$ (as in natural movies) (Fig. 5e). Thus, as the minimal angle $\alpha_{\min}$ increases, overall adaptation modulation decreases, indicating weaker adaptation when neural responses are more separated, as in the natural movies case. A reduced exponent magnitude thereby signifies a weakened adaptation process. Our approach, which models adaptation to one-dimensional stimuli such as orientations, is thus able to reproduce the power law for oriented, one-dimensional stimuli as well as the power law of the more realistic natural movies (Fig. 5f), by assuming that the only property that matters in determining the power law's exponent is the spacing between neural responses.

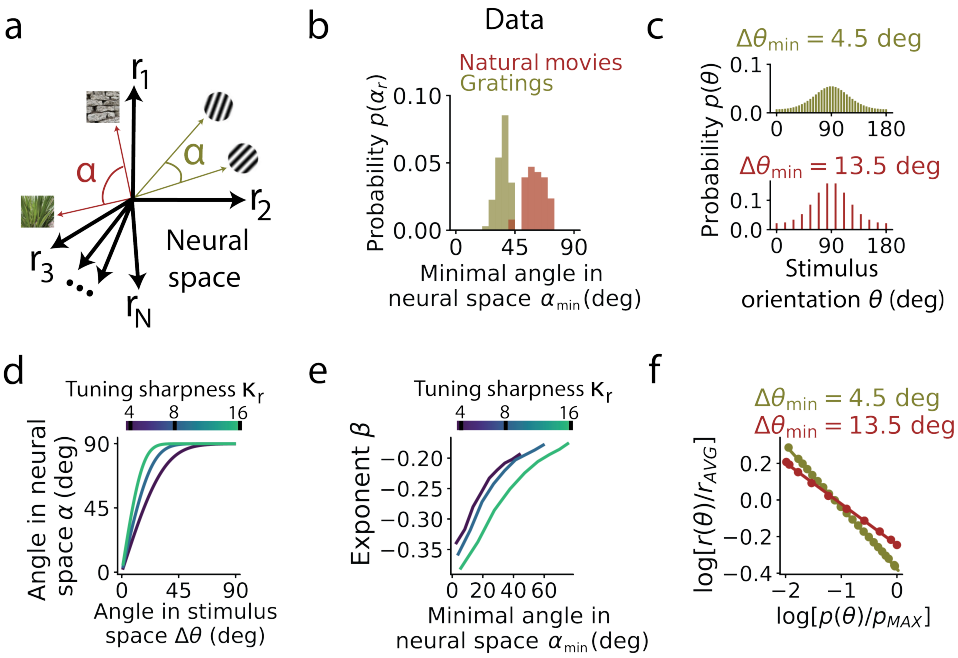

Figure 5: **The spacing between neural representations explains the change in power law exponent in natural movies.** a) Outline of the angle between representations in the high dimensional neural space identified as response vectors to a pair of oriented gratings (yellow) and a pair of natural stimuli (red). b) Histogram of minimal angular distance between neural responses to a finite number of nearby oriented gratings (yellow) and natural movies (red), obtained by analyzing data from [1]. c) Two examples of spaced environments, namely discrete environments with a fixed minimal distance $\Delta\theta_{\min}$ between stimuli, which is narrower in the top panel (yellow) and larger in the bottom panel (red). d) Angle between neural representations α as a function of angle between stimuli $\Delta\theta$, for three different values of tuning curves sharpness κ_r, marked with black ticks in the color bar. e) The power law exponent β as a function of the minimal angle $\alpha_{\min}$ between response vectors, for the values of κ_r in (d). e) The logarithm of the response rate as a function of the logarithm of the stimulus probability, for the two environments shown in panel (c). The probability is normalized by the maximum value while the rate is normalized respectively by the average value.

Discussion

The discovery of the power law in [1], which links the probability of stimuli in an environment to the adapted average population response of sensory neurons, enables precise predictions about how neural responses adapt to new environments. In this context, we introduced a model for adaptation that, for the first time, offers a normative explanation for the observed power law. We postulated that the adapted tuning curves to a given statistical environment can be found by minimizing a multi-objective function consisting of a weighted sum of an energy cost, a detection error, a discrimination error, a homeostasis constraint and a smoothness constraint. We found that the power-law exponent predicted by our model is invariant to changes in the environment, consistent with experimental findings. We interpreted the exponent as reflecting the trade-off between energy and representation fidelity in adaptation. We found that the observed change in exponent for natural images results from an increased minimal angle between response vectors. Thus, our work provides a simple, interpretable, and testable model that explains the functional benefits of the power law and its related properties discovered in [1].

We validated our hypotheses against experimental data. Reproducing the universal power law served as an initial confirmation of our model. We also performed a new analysis of the data in [1] and discovered that the average modulation of the tuning curves during adaptation is primarily stimulus-dependent, a feature our model successfully captures. By adjusting the five parameters that weight the costs, we also showed that the model's average population responses align with the experimental data. Using these tuned parameters, we then predicted the adapted response rates in new environments. We were able to explain the magnitude reduction of the power law exponent observed in responses to natural movies (Figure 5). We reproduced this effect using one-dimensional stimuli, ensuring that consecutive stimuli were sufficiently far apart. This resulted in response vectors that were widely separated in angle, similar to those observed in natural movie data. As the angle between response vectors increased, we observed a smooth transition from the gratings exponent to the natural movies exponent. Collectively, these validations confirmed the robustness of our model.

The model is based on the efficient coding approach [39], [40], [45], namely that sensory neurons optimize the representation of the environment given a limited amount of available resources. Although it is widely accepted that neural processing is shaped by environmental statistics, quantitatively defining this relationship remains challenging [46]. Various theories have been proposed, such as sensory neurons maximizing mutual information between stimuli and their representations under limited coding capacity [47]–[50], or aiming to eliminate redundancy in representations. Other perspectives suggest that the brain seeks to infer high-level hidden features of the environment within the

constraints of a limited entropy rate [33], or that it minimizes the average reconstruction error while reducing metabolic costs [38]. The optimization process may also account for the role of downstream motor neurons [51] in the optimization process. The selection of a specific (multi)objective function within the efficient coding framework therefore depends on the particular phenomenon being modeled [46], [48]. The negative power-law exponent observed in the data suggested that energy conservation plays a key role in adaptation. Consequently, we introduced the energy cost term. The other cost terms were designed around this principle. However, our approach does not rule out the possibility that other frameworks could also explain the power law. Whether alternative paradigms can account for the full set of experimental results of [1], as our model does, remains an open question.

Our model is highly interpretable and we provided insight into the role of each term of the multi-objective function. The energy cost, detection error and homeostasis constraint are necessary and sufficient to get a negative power law. Furthermore, the discrimination error is necessary to explain the tilt of response vectors caused by adaptation. Finally, the smoothness constraint is necessary to explain the continuity of average population responses in discontinuous environments.

We observed that the exponent is a decreasing function of the energy weight λ_E, transitioning from positive to negative values as λ_E increases when the other weights are fixed (Figure 3). Similar exponent changes can be achieved by keeping the energy weight λ_E fixed while varying the other weights. Animals may dynamically adjust the relative magnitude of the weights, and thus the power law exponent, based on their energy-saving needs. This could explain changes in neural responses during locomotion [52], [53].

Our approach relies on several hypotheses. We parameterized the average population responses of neurons using tuning curves, and hypothesized that the noise in neural representations is Gaussian, independent and isotropic. We also assumed that adaptation affects the tuning curves only, without modifying the noise distribution. Further theoretical development may relax the latter hypothesis and test other distributions of noise, such as Poisson. An important next step is to extend the model to additional stimulus features. For example, [10] examined contrast variation across stimuli and identified a power law linking responses to contrast.

Our model generates predictions about adapted tuning curves, overall firing rates, and discrimination and detection errors, that could guide future research. For example, future experiments may validate or refute our model's prediction that an increased minimum angle between neural representations leads to a change in the exponent (Figure 5e). One possible direction would be to investigate two-dimensional stimuli, such as gratings varying in both angle and spatial frequency, where we expect a greater minimal distance between a finite number of neural representations and a corresponding exponent closer to zero than in the one-dimensional case.

Acknowledgments

This work was supported by NIH R21 EY035064 (DLR, MD, and MM). We thank Ramon Nogueira and Dean Buonomano for helpful comments on this manuscript.

Author contributions

DLR, MD, and MM designed the model; MD and MM developed the model; MD and MM designed the data analysis; MM performed the data analysis; DLR, MD, MM, and SAM, interpreted results; MD and MM wrote the original draft; DLR, MD, MM, and SAM edited the manuscript.

Methods

Sensory environments Each stimulus was represented by an angle θ in the interval $[0, \pi]$, corresponding to the orientation of a grating (in this context, an angle $\theta + \pi$ is equivalent to θ). A sensory environment e was defined by the probability density function (PDF) $p_e(\theta)$ that the stimulus was presented. We studied continuous environments, namely the ones having a continuous PDF (Suppl. Figure S.6a). Additionally, we introduced environments defined on $[0, \pi]$ with a finite minimal distance $\Delta\theta$ between any pair of stimuli (Suppl. Figure S.6b). We refer to these as *spaced environments* and use them to model the representation of natural movies in neural data (Figure 5). To enforce a minimal distance between equally spaced stimuli, only a finite set can have nonzero probability. Consequently, the probability density function (PDF) of a spaced environment is expressed as a sum of Dirac delta functions $\delta(\theta - k\Delta\theta)$, each corresponding to a stimulus k and weighted by its probability p_k. The spaced environment PDF is therefore: $p_e(\theta) = \sum_{k=1}^{n} p_k \delta(\theta - k\Delta\theta)$, where n is the total number of stimuli and $\Delta\theta = \pi/n$ is the spacing between two consecutive stimuli, which is given by the ratio of the domain width, π, to the total number of stimuli in the domain. Regardless of the environment studied, continuous or spaced, we discretized for computational reasons the domain $[0, \pi]$ in n_S values. As a remark, the total number n of discrete stimuli is different from the total number n_S of discretization steps of the domain, with $n_S > n$. In the discrete case, only n out of n_S values of the domain have a non zero probability (Suppl. Figure S.6b).

Tuning curves Our model considered reference tuning curves $r_0(\varphi, \theta)$ defined over the stimulus orientation θ, of neurons with preferred orientation φ. The reference tuning curves were modeled by von Mises functions:

$$r_0(\varphi, \theta) = R_0 \frac{\exp\{\kappa_r \cos[2(\theta - \varphi)]\}}{2\pi I_0(\kappa_r)}, \tag{3}$$

where κ_r measures the sharpness of the tuning curves and $I_0(x)$ is the modified Bessel function of the first kind of order 0. The factor 2 in front of the angular terms in the r.h.s. of the equation ensures that the function is periodic in $[0, \pi]$ instead of $[0, 2\pi]$. The prefactor R_0 is a number that represents the reference amplitude of the responses. This quantity needed to be fixed as a hyperparameter (Supplementary Fig. S.4) and could not be obtained by the optimization process of the total cost in eq. (2). However, as long as the initial amplitude R_0 is sufficiently large, the results of the model are independent of the specific value of R_0 (Supplementary fig. S.4). Increasing the value of R_0 ensures the emergence of a power law.

Multi-objective optimization For a given environment $p_e(\theta)$, the model finds the neuron-dependent and stimulus-dependent modulations, respectively $a_1(\varphi)$ and $a_2(\theta)$, such that the modulated tuning curves $r(\varphi, \theta)$ minimize the total cost in equation (2). The energy cost was defined as:

$$\mathcal{E}[p_e(\theta), r(\varphi, \theta)] = \mathbb{E}_{\theta \sim p_e(\theta)}[\|r(\cdot, \theta)\|^2_{L_2}], \tag{4}$$

where the norm $\|r(\cdot, \theta)\|_{L_2}$ was defined as the square of the L_2 norm over the neurons:

$$\|r(\cdot, \theta)\|^2_{L_2} = \frac{1}{\pi} \int_0^\pi r(\varphi, \theta)^2 \mathrm{d}\varphi$$

This quantity is related to the average population response over time, according to the idea that spiking consumes energy. The symbol $\mathbb{E}$ is the expected value. The subscript $\theta \sim p_e(\theta)$ specifies that the average is with respect to the angular stimulus θ, which is stochastically drawn from the environment distribution of stimuli $p_e(\theta)$:

$$\mathbb{E}_{\theta \sim p_e(\theta)}[f(\theta)] = \int_0^\pi p_e(\theta) f(\theta) \mathrm{d}\theta$$

for a generic function f. The discrimination error was defined starting from the definition of d-prime [54]–[56]. As a simplifying assumption, we modeled the noise as independent, normal, and identically distributed across neurons, with a covariance matrix equal to $\mathbb{I}\sigma^2$. For an independent, normal and identically distributed noise, the d-prime between the responses to two stimuli θ_1 and θ_2 is:

$$d'(\theta_1, \theta_2) = \frac{1}{\sigma} \|r(\cdot, \theta_1) - r(\cdot, \theta_2)\|_{L_2},$$

which can be interpreted as a signal-to-noise ratio, where the signal is the difference between mean responses. We defined the discrimination error by

$$\mathcal{D}[p_e(\theta), r(\varphi, \theta)] = 1 - \mathbb{E}_{\theta_1, \theta_2 \sim p_e(\theta)}[\Phi\left(d'(\theta_1, \theta_2)\right)] = \tag{5}$$

$$= 1 - \mathbb{E}_{\theta_1, \theta_2 \sim p_e(\theta)}[\Phi\left(\|r(\cdot, \theta_1) - r(\cdot, \theta_2)\|_{L_2}/\sigma\right)], \tag{6}$$

where the function $\Phi(z)$ is the cumulative distribution function for a normal distribution of zero mean and unitary variance [54]:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2}, \tag{7}$$

and where we averaged over pairs of stimuli, θ_1 and θ_2, drawn independently from the environment distribution $p_e(\theta)$. The average over θ_1 and θ_2 should be interpreted as a time average over couples of consecutive stimuli that the brain is trying to discriminate, and the cost defined thus represents the average discrimination error in the task.

The detection error was defined as:

$$\mathcal{T}[p_e(\theta), r(\varphi, \theta)] = 1 - \mathbb{E}_{\theta \sim p_e(\theta)}\left[\Phi\left(\|r(\cdot, \theta)\|_{L_2}\right)\right]. \tag{8}$$

This term measures the error in the detection of a stimulus against a background. It is equivalent to the average error in the discrimination between the stimulus presented and another with zero contrast, which we model to elicit a response with zero amplitude.

The homeostasis constraint was defined as:

$$\mathcal{H}[r(\varphi, \theta)] = \frac{1}{\pi} \int_0^\pi \|r(\cdot, \theta) - r_0(\cdot, \theta)\|^2_{L_2} \mathrm{d}\theta, \tag{9}$$

where $r_0(\varphi, \theta)$ is the reference tuning curve defined in equation (3). This term penalizes adapted tuning curves that are too different from the reference ones [57]. Note that this form of homeostasis is different from the one described

as a cause of sensory adaptation [58], which is instead defined as the invariance of the average firing rates across the population of neurons and across environments.

The smoothness constraint was defined as:

$$\mathcal{S}[r(\varphi,\theta)] = \frac{1}{\pi}\int_0^{\pi}\int_0^{\pi}\left[\frac{\mathrm{d}r(\varphi,\theta)}{\mathrm{d}\theta}\right]^4 \mathrm{d}\varphi\mathrm{d}\theta. \tag{10}$$

The smoothness term ensures that neurons do not change abruptly their responses to similar stimuli. The fourth power is to penalize discontinuities, while neglecting small changes. A lower exponent in the smoothness constraint was not strong enough to prevent the tuning curves to change abruptly when the probability distribution changes in a discontinuous way. Thus, if we used a lower exponent, we would have needed to increase the smoothness weight λ_S to make the smoothness constraint sufficiently strong. However, in that case, the smoothness constraint would have always been dominant with respect to the other costs regardless of the environment. In the case of a quadratic term, the smoothness constraint would induce average population responses too close to a constant, which is not compatible with the data.

The optimization process We minimized the total cost $\mathcal{C}$ defined in eq. (2) using a gradient descent technique. We enforced a nonparametric optimization, discretizing the orientation space into n_S stimuli and the neural space into n_R neurons. The modulation functions, a_1 and a_2, were discretized as $a_1 = \{a_1(\varphi_1), a_1(\varphi_2), \ldots, a_1(\varphi_{n_R})\}$ and $\{a_2 = (a_2(\theta_1), a_2(\theta_2), \ldots, a_2(\theta_{n_S})\}$. The number of optimized parameters was thus $n_S + n_R$. We initialized these terms to 1 and subsequently updated them via a gradient descent over the total cost. The gradient descent procedure was structured as follows:

- We computed $r(\varphi,\theta) = r_0(\varphi,\theta)a_1(\varphi)a_2(\theta)$ for each $\varphi \in \{\varphi_1, \ldots, \varphi_{n_R}\}$ and each $\theta \in \{\theta_1, \ldots, \theta_{n_S}\}$
- We substituted the function $r(\varphi,\theta)$ onto the total cost $\mathcal{C}$ in equation (2) and we computed an auto-gradient numerically with respect to the $n_R + n_S$ parameters. In the discretized form each cost became:

$$\begin{cases} \mathcal{E} = \sum_{i=1}^{n_S}\frac{1}{n_R}\sum_{j=1}^{n_R} r(\varphi_j,\theta_i)^2 p(\theta_i) \\ \mathcal{D} = \sum_{i_1=1}^{n_S}\sum_{i_2=1}^{n_S}\Phi\left(\sqrt{\frac{1}{n_R}\sum_{j=1}^{n_R}(r(\varphi_j,\theta_{i_1}) - r(\varphi_j,\theta_{i_2}))^2}\right) p(\theta_{i_1})p(\theta_{i_2}) \\ \mathcal{T} = \sum_{i=1}^{n_S}\Phi\left(\sqrt{\frac{1}{n_R}\sum_{j=1}^{n_R} r(\varphi_j,\theta_i)^2}\right) p(\theta_i) \\ \mathcal{H} = \pi\frac{1}{n_S}\sum_{i=1}^{n_S}\frac{1}{n_R}\sum_{j=1}^{n_R}\left[r(\varphi_j,\theta_i) - r_0(\varphi_j,\theta_i)\right]^2 \\ \mathcal{S} = \pi n_S^3\sum_{i=1}^{n_S}\frac{1}{n_R}\sum_{j=1}^{n_R}\left[r(\varphi_j,\theta_{i+1}) - r(\varphi_j,\theta_i)\right]^4. \end{cases} \tag{11}$$

  We have set the noise to a reference value of one, $\sigma^2 = 1$. This is an arbitrary choice, which does not reduce generality, since the noise only gives a scale for the intensity of rates. By setting $\sigma^2 = 1$, we are fixing such a scale.

- We updated each parameter by subtracting along the sign of the partial derivative, using a gradient descent method named Sign Descender [59], an effective method in numerical optimization problems [60].
- We repeated the procedure until convergence.

Once the optimal tuning curves were found, we computed the average population response as a function of the stimuli: $r(\theta) = \frac{1}{n_R}\sum_\varphi r(\varphi,\theta)$. As a last step, we plotted $\log r(\theta)$ as a function of $\log p(\theta)$ and we linearly fitted $\log r(\theta)$ as a function of $\log p(\theta)$: $\log r(\theta) = \beta \log p(\theta) + q$. This linear relation implies the power law: $r(\theta) = e^q p(\theta)^\beta \propto p(\theta)^\beta$. Thus, by evaluating the score of the fit, it was possible to evaluate how close the relation was to a power law.

Training the model to study the modulation of single tuning curves

We studied whether our model could reproduce a power law by modulating inhomogeneous tuning curves matching the experimental data [1]. To achieve this goal, we first optimized the weights $\vec{\lambda}$ to reproduce the recorded neural responses in a specific environment. Specifically, we selected a training environment A out of the three environments in the experimental data. We initialized the weight vector to the one used in the study of the power law in synthetic environments and synthetic reference tuning curves (Figures 2 and 3). We denote the recorded tuning curves from [1]

by $t(\varphi, \theta)$. We constructed the reference tuning curves $r_0(\varphi, \theta)$ based on the recorded tuning curves in the training environment $t_A(\varphi, \theta)$, normalized by their stimulus:

$$r_0(\varphi, \theta_k) = \frac{t_A(\varphi, \theta_k)}{\sum_i t_A(\varphi, \theta_i)}.$$

Note that the form of the reference tuning curves $r_0(\varphi, \theta)$ is an arbitrary choice in the model. The model works for several choices of reference tuning curves. We opted to choose a form that retained as much as possible the information in the training data. At fixed $\vec{\lambda}$, we minimized the total cost in equation (2) and obtained the adapted rates predicted by our model, $r_A(\varphi, \theta; \vec{\lambda})$, where the dependence on the weight vector $\vec{\lambda}$ is explicitly shown. We then updated the weight vector $\vec{\lambda}$ by descending the gradient of the error between the model's tuning curves and the recorded tuning curves:

$$\mathrm{Err}(\vec{\lambda}) = \sum_{\varphi,\theta} \left[r_A(\varphi, \theta; \vec{\lambda}) - t_A(\varphi, \theta) \right]^2. \tag{12}$$

Specifically, we evaluated the gradient numerically, via finite difference, and used it to update $\vec{\lambda}$ using the Sign Descender method [59]. We used the updated weight vector to compute the tuning curves optimizing the cost. We stopped the procedure when both the weight vector and the error in the tuning curves converged. The optimal $\vec{\lambda}^*$, the one minimizing the error, was then used to predict the average population responses in either of the other two environments, which we compared to the recorded responses from the experiments in [1] (Figure 4d-e).

Experimental analyses

We analyzed data from [1], which consisted of 20 recording sessions from four adult mice in the mouse primary visual cortex (V1). In each recording session, a head-fixed mouse passively viewed oriented gratings while running on a treadmill. Neural responses from hundreds of V1 neurons were simultaneously recorded using calcium imaging. The stimuli were sets of 18 oriented gratings or 18 natural movies, each displayed on a screen for 300 ms. Each recording session included six blocks of approximately 100 stimuli. Within a block, stimuli were drawn from three smaller blocks randomly permuted, each one corresponding to a different environment.

Analysis of modulation in experimental data To determine the structure of the adaptation-induced modulation (Fig. 2e), we considered only recording sessions where one of the environments was uniform. We then used these recordings to compute the non-adapted average tuning curves of neurons with similar preferred orientations. We reduced the original set of neurons to a set of $n_R = 18$ super-neurons [38], [58], [61]. These super neurons are defined by taking the average of all neurons whose preferred orientation was within a given range of orientations. The orientations defining the borders of these ranges were evenly spaced within $[0, \pi]$. We computed the modulation by taking the ratio of the super-neurons tuning curves, $a(\varphi, \theta) = t_e(\varphi, \theta)/t_U(\varphi, \theta)$, where U stands for the uniform environment, $e \in \{A, B\}$ is one of the two biased environments and $t_x(\varphi, \theta)$ are the recorded tuning curves for environment x. All the biased environments of these selected recording sessions followed a von Mises distribution with the same sharpness, differing only in the position of the peak. This symmetry allowed us to average the modulation across biased environments and recording sessions by applying a circular shift to the modulation to align each one with the same reference stimulus and preferred orientation.

Angular spacing between responses We studied the angle between response vectors to different stimuli, to compare the results for oriented gratings to those for natural movies. Let us consider the response vector $\mathbf{r}(x) = \{r_1(x), r_2(x), \ldots, r_{n_R}(x)\}$ to a stimulus x, which could represent either an orientated grating or a natural movie. Each element of the response vector corresponds to a different neuron rate among the n_R neurons recorded. We computed the angle between the response vectors to two different stimuli x_1 and x_2 as:

$$\alpha(x_1, x_2) = \arccos\left[\frac{\mathbf{r}(x_1) \cdot \mathbf{r}(x_2)}{\|\mathbf{r}(x_1)\| \|\mathbf{r}(x_2)\|} \right]. \tag{13}$$

We used such a computation in the study of the difference in exponent between gratings and natural movies (Figure 5). In particular, we characterized the distribution of minimal angles between response vectors to natural movies and oriented gratings. To do so, we considered, for each stimulus x_1, the stimulus x_2 such that the angle $\alpha(x_1, x_2)$ was minimal. We iterated through the three recording sessions of natural movies in [1], and for each of them we iterated through the three environments recorded. We built the histogram of all $\alpha_{\min}(x_1, x_2)$ across recording sessions and environments in Figure 5b.

References

[1] Tring, E., Dipoppa, M., and Ringach, D. L., A power law describes the magnitude of adaptation in neural populations of primary visual cortex, *Nature Communications*, vol. 14, 8366, 2023.

[2] Westrick, Z. M., Heeger, D. J., and Landy, M. S., Pattern Adaptation and Normalization Reweighting, *Journal of Neuroscience*, vol. 36, 9805–9816, 2016.

[3] Clifford, C. W. G., Webster, M. A., Stanley, G. B., *et al.*, Visual adaptation: Neural, psychological and computational aspects, *Vision Research*, vol. 47, 3125–3131, 2007.

[4] Webster, M. A., Adaptation and visual coding, *Journal of Vision*, vol. 11, 3, 2011.

[5] Whitmire, C. J. and Stanley, G. B., Rapid Sensory Adaptation Redux: A Circuit Perspective, *Neuron*, vol. 92, 298–315, 2016.

[6] Borst, A., Flanagin, V. L., and Sompolinsky, H., Adaptation without parameter change: Dynamic gain control in motion detection, *Proceedings of the National Academy of Sciences*, vol. 102, 6172–6176, 2005.

[7] Kohn, A., Visual Adaptation: Physiology, Mechanisms, and Functional Benefits, *Journal of Neurophysiology*, vol. 97, 3155–3164, 2007.

[8] Brown, S. P. and Masland, R. H., Spatial scale and cellular substrate of contrast adaptation by retinal ganglion cells, *Nature Neuroscience*, vol. 4, 44–51, 2001.

[9] Kastner, D. B. and Baccus, S. A., Insights from the retina into the diverse and general computations of adaptation, detection, and prediction, *Current Opinion in Neurobiology*, vol. 25, 63–69, 2014.

[10] Tring, E., Dipoppa, M., and Ringach, D. L., On the contrast response function of adapted neural populations, *Journal of Neurophysiology*, vol. 131, 446–453, 2024.

[11] Crowder, N. A., Price, N., Hietanen, M. A., Dreher, B., Clifford, C., and Ibbotson, M. R., Relationship Between Contrast Adaptation and Orientation Tuning in V1 and V2 of Cat Visual Cortex, *Journal of Neurophysiology*, vol. 95, 271–283, 2006.

[12] Patterson, C. A., Wissig, S. C., and Kohn, A., Distinct Effects of Brief and Prolonged Adaptation on Orientation Tuning in Primary Visual Cortex, *Journal of Neuroscience*, vol. 33, 532–543, 2013.

[13] Tolias, A. S., Keliris, G. A., Smirnakis, S. M., and Logothetis, N. K., Neurons in macaque area V4 acquire directional tuning after adaptation to motion stimuli, *Nature Neuroscience*, vol. 8, 591–593, 2005.

[14] Krekelberg, B., Wezel, R. J. A. van, and Albright, T. D., Adaptation in Macaque MT Reduces Perceived Speed and Improves Speed Discrimination, *Journal of Neurophysiology*, vol. 95, 255–270, 2006.

[15] Fairhall, A. L., Lewen, G. D., Bialek, W., and Ruyter van Steveninck, R. R. de, Efficiency and ambiguity in an adaptive neural code, *Nature*, vol. 412, 787–792, 2001.

[16] Baccus, S. A. and Meister, M., Fast and Slow Contrast Adaptation in Retinal Circuitry, *Neuron*, vol. 36, 909–919, 2002.

[17] Carandini, M., Anthony Movshon, J., and Ferster, D., Pattern adaptation and cross-orientation interactions in the primary visual cortex, *Neuropharmacology*, vol. 37, 501–511, 1998.

[18] Carandini, M., Barlow, H. B., O'Keefe, L. P., Poirson, A. B., and Movshon, J. A., Adaptation to contingencies in macaque primary visual cortex, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 352, 1149–1154, 1997.

[19] Dragoi, V., Sharma, J., and Sur, M., Adaptation-Induced Plasticity of Orientation Tuning in Adult Visual Cortex, *Neuron*, vol. 28, 287–298, 2000.

[20] Clifford, C. W. G., The tilt illusion: Phenomenology and functional implications, *Vision Research*, vol. 104, 3–11, 2014.

[21] Schwartz, O., Hsu, A., and Dayan, P., Space and time in visual context, *Nature Reviews Neuroscience*, vol. 8, 522–535, 2007.

[22] Kloth, N., Schweinberger, S. R., and Kovács, G., Neural Correlates of Generic versus Gender-specific Face Adaptation, *Journal of Cognitive Neuroscience*, vol. 22, 2345–2356, 2010.

[23] Weber, A. I., Krishnamurthy, K., and Fairhall, A. L., Coding Principles in Adaptation, *Annual Review of Vision Science*, vol. 5, 427–449, 2019.

[24] Brinkman, B. A. W., Weber, A. I., Rieke, F., and Shea-Brown, E., How Do Efficient Coding Strategies Depend on Origins of Noise in Neural Circuits? *PLOS Computational Biology*, vol. 12, e1005150, 2016.

[25] Duong, L., Simoncelli, E., Chklovskii, D., and Lipshutz, D., Adaptive whitening with fast gain modulation and slow synaptic plasticity, *Advances in Neural Information Processing Systems*, vol. 36, 12518–12535, 2023.

[26] Rao, R. P. N. and Ballard, D. H., Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects, *Nature Neuroscience*, vol. 2, 79–87, 1999.

[27] Keller, G. B. and Mrsic-Flogel, T. D., Predictive Processing: A Canonical Cortical Computation, *Neuron*, vol. 100, 424–435, 2018.

[28] Spratling, M. W., A review of predictive coding algorithms, *Brain and Cognition*, vol. 112, 92–97, 2017.

[29] Mao, J., Rothkopf, C. A., and Stocker, A. A., Adaptation optimizes sensory encoding for future stimuli, *PLOS Computational Biology*, vol. 21, e1012746, 2025.

[30] Ulanovsky, N., Las, L., and Nelken, I., Processing of low-probability sounds by cortical neurons, *Nature Neuroscience*, vol. 6, 391–398, 2003.

[31] Hosoya, T., Baccus, S. A., and Meister, M., Dynamic predictive coding by the retina, *Nature*, vol. 436, 71–77, 2005.

[32] Kersten, D., Mamassian, P., and Yuille, A., Object Perception as Bayesian Inference, *Annual Review of Psychology*, vol. 55, 271–304, 2004.

[33] Młynarski, W. F. and Hermundstad, A. M., Adaptive coding for dynamic sensory inference, *eLife*, vol. 7, e32055, 2018.

[34] Wainwright, M. J., Visual adaptation as optimal information transmission, *Vision Research*, vol. 39, 3960–3974, 1999.

[35] Brenner, N., Bialek, W., and Ruyter van Steveninck, R. de, Adaptive Rescaling Maximizes Information Transmission, *Neuron*, vol. 26, 695–702, 2000.

[36] Sharpee, T. O., Calhoun, A. J., and Chalasani, S. H., Information theory of adaptation in neurons, behavior, and mood, *Current Opinion in Neurobiology*, vol. 25, 47–53, 2014.

[37] Młynarski, W. F. and Hermundstad, A. M., Efficient and adaptive sensory codes, *Nature Neuroscience*, vol. 24, 998–1009, 2021.

[38] Dipoppa, M., Nogueira, R., Bugeon, S., *et al.*, Adaptation shapes the representational geometry in mouse V1 to efficiently encode the environment, *bioRxiv*, 628035, 2024.

[39] Attneave, F., Some informational aspects of visual perception, *Psychological Review*, vol. 61, 183–193, 1954.

[40] Barlow, H., Possible Principles Underlying the Transformations of Sensory Messages, *Sensory Communication*, vol. 1, 1961.

[41] Linsker, R., Self-Organization in a Perceptual Network. *IEEE Computer*, vol. 21, 105–117, 1988.

[42] Ganguli, D. and Simoncelli, E. P., Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations, *Neural Computation*, vol. 26, 2103–2134, 2014.

[43] Gutierrez, G. J. and Denève, S., Population adaptation in efficient balanced networks, *eLife*, vol. 8, e46926, 2019.

[44] Li, J. Y. and Glickfeld, L. L., Input-specific synaptic depression shapes temporal integration in mouse visual cortex, *Neuron*, vol. 111, 3255–3269.e6, 2023.

[45] Miller, G. A., What is information measurement? *American Psychologist*, vol. 8, 3–11, 1953.

[46] Simoncelli, E. P. and Olshausen, B. A., Natural Image Statistics and Neural Representation, *Annual Review of Neuroscience*, vol. 24, 1193–1216, 2001.

[47] Wei, X.-x. and Stocker, A. A., Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference, in *Advances in Neural Information Processing Systems*, vol. 25, 2012.

[48] Wei, X.-X. and Stocker, A. A., A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts, *Nature Neuroscience*, vol. 18, 1509–1517, 2015.

[49] Morais, M. and Pillow, J. W., Power-law efficient neural codes provide general link between perceptual bias and discriminability, in *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[50] Hahn, M. and Wei, X.-X., A unifying theory explains seemingly contradictory biases in perceptual estimation, *Nature Neuroscience*, vol. 27, 793–804, 2024.

[51] Salinas, E., How Behavioral Constraints May Determine Optimal Sensory Representations, *PLOS Biology*, vol. 4, e387, 2006.

[52] Niell, C. M. and Stryker, M. P., Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex, *Neuron*, vol. 65, 472–479, 2010.

[53] Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., and Harris, K. D., Vision and Locomotion Shape the Interactions between Neuron Types in Mouse Visual Cortex, *Neuron*, vol. 98, 602–615.e8, 2018.

[54] Nogueira, R., Peltier, N. E., Anzai, A., DeAngelis, G. C., Martínez-Trujillo, J., and Moreno-Bote, R., The Effects of Population Tuning and Trial-by-Trial Variability on Information Encoding and Behavior, *Journal of Neuroscience*, vol. 40, 1066–1083, 2020.

[55] Averbeck, B. B. and Lee, D., Effects of Noise Correlations on Information Encoding and Decoding, *Journal of Neurophysiology*, vol. 95, 3633–3644, 2006.

[56] Chen, Y., Geisler, W. S., and Seidemann, E., Optimal decoding of correlated neural population responses in the primate visual cortex, *Nature Neuroscience*, vol. 9, 1412–1420, 2006.

[57] Duong, L. R., Bredenberg, C., Heeger, D. J., and Simoncelli, E. P., Adaptive coding efficiency in recurrent cortical circuits via gain control, *arXiv*, 19869, 2023.

[58] Benucci, A., Saleem, A. B., and Carandini, M., Adaptation maintains population homeostasis in primary visual cortex, *Nature Neuroscience*, vol. 16, 724–729, 2013.

[59] Safaryan, M. and Richtarik, P., Stochastic Sign Descent Methods: New Algorithms and Better Theory, in *Proceedings of the 38th International Conference on Machine Learning*, 2021, 9224–9234.

[60] Braunstein, A., Catania, G., Dall'Asta, L., Mariani, M., and Muntoni, A. P., Inference in conditioned dynamics through causality restoration, *Scientific Reports*, vol. 13, 7350, 2023.

[61] Stringer, C., Michaelos, M., Tsyboulski, D., Lindo, S. E., and Pachitariu, M., High-precision coding in visual cortex, *Cell*, vol. 184, 2767–2778.e15, 2021.

Supplementary information

Analytic approach for a simplified model

We considered a simpler optimization problem than the one considered in equation (2) and we solved it analytically, showing that a power law emerges across a broad region of parameter space. This simplified model optimizes only energy cost, detection error, and homeostasis constraint:

$$\mathcal{C} = \lambda_E \mathcal{E} + \lambda_T \mathcal{T} + \lambda_H \mathcal{H}. \tag{S.1}$$

To analytically tackle the optimization of this function, we first introduced two simplifying hypotheses:

1. In the detection error, defined in equation(8), we substituted $\Phi(z)$ with the identity z, so that the detection error became:
$$\mathcal{T}[r] = 1 - \mathbb{E}_{\theta\sim p(\theta)}\left[\|\mathbf{r}(\theta)\|\right]. \tag{S.2}$$
Since the noise is set to the reference unitary value , $\sigma^2 = 1$, this formula corresponds to one minus the average signal-to-noise ratio of responses.

2. We imposed that the modulation is only dependent on the stimulus:
$$r(\varphi, \theta) = a(\theta) r_0(\varphi, \theta) \tag{S.3}$$
where $r_0(\varphi, \theta)$ are the reference tuning curves, and $a(\theta)$ is their modulation due to adaptation. The dependence of the modulation on the stimulus alone is an approximation, justified by the fact that the modulation in the full model is mainly dependent on the stimulus (Figure 2d-e).

We considered a discretized form of the total cost to compute its optimization by means of partial derivatives rather then functional derivative. We discretized the continuous stimulus space into n_S values, $\theta \in \{\theta_1, \theta_2, \ldots, \theta_{n_S}\}$. The probability of stimuli thus becomes $p(\theta) = \{p(\theta_1), p(\theta_2), \ldots, p(\theta_{n_S})\}$, and the modulation becomes $a(\theta) = \{(a(\theta_1), a(\theta_2), \ldots, a(\theta_{n_S})\}$. Therefore, the number of stimuli n_S represents the total number of discretization steps. The total cost for the simplified model is thus:

$$C = \lambda_E \sum_{i=1}^{n_S}\sum_{j=1}^{n_R} p(\theta_i)\frac{r(\varphi_j,\theta_i)^2}{n_R} + \lambda_T\left[1 - \sum_{i=1}^{n_S} p(\theta_i)\sqrt{\frac{1}{n_R}\sum_{j=1}^{n_R} r(\varphi_j,\theta_i)^2}\right] + \lambda_H \frac{\pi}{n_S}\sum_{i=1}^{n_S}\frac{1}{n_R}\sum_{j=1}^{n_R}\left(r(\varphi_j,\theta_i) - r_0(\varphi_j,\theta_i)\right)^2$$

where we have also discretized the number of neurons into n_R values $\varphi \in \{\varphi_1, \varphi_2, \ldots, \varphi_{n_R}\}$. Using equation (S.3), the total cost becomes:

$$C = \lambda_E \sum_{i,j} p(\theta_i) a(\theta_i)^2 \frac{r_0(\varphi_j,\theta_i)^2}{n_R} + \lambda_T\left[1 - \sum_i p(\theta_i) a(\theta_i)\sqrt{\frac{1}{n_R}\sum_j r_0(\varphi_j,\theta_i)^2}\right] + \frac{\pi\lambda_H}{n_R n_S}\sum_{i,j} r_0^2(\varphi_j,\theta_i)\left(a(\theta_i) - 1\right)^2. \tag{S.4}$$

We define now:

$$F_0^2 = \frac{1}{n_R}\sum_j r_0^2(\varphi_j, \theta),$$

which is independent of θ for the reference tuning curves. Note that F_0 is approximately proportional to the amplitude of the reference tuning curve R_0 when the number of neurons n_R is large:

$$F_0^2 = \frac{R_0^2}{n_R}\sum_j \frac{\exp(2\kappa_r\cos(\theta - \varphi_j))}{2\pi I_0(\kappa_r)}. \tag{S.5}$$

Defining $\mathrm{d}\varphi = \pi/n_R$:

$$F_0^2 \approx \frac{R_0^2}{\pi}\int \frac{\exp(2\kappa_r\cos(\theta - \varphi))}{2\pi I_0(\kappa_r)}\mathrm{d}\varphi = \frac{R_0^2}{\pi}\frac{I_0(2\kappa_r)}{I_0(\kappa_r)}. \tag{S.6}$$

Inserting F_0 in the total cost, eq. (S.4), we obtain:

$$C = \lambda_E \sum_i a(\theta_i)^2 F_0^2 p(\theta_i) - \lambda_T \sum_i a(\theta_i) F_0 p(\theta_i) + \frac{\pi\lambda_H}{n_S}\sum_i F_0^2\left(a(\theta_i) - 1\right)^2. \tag{S.7}$$

To find the local minima of the total cost of eq. (S.7), we set to zero the partial derivative with respect to a generic element $a(\theta)$, with $\theta \in \{\theta_1, \ldots, \theta_{n_S}\}$:

$$0 = \frac{\partial\mathcal{C}}{\partial a(\theta)} = \lambda_E 2a(\theta)F_0^2 p(\theta) - \lambda_T F_0 p(\theta) + \frac{\pi\lambda_H}{n_S}F_0^2 2\left(a(\theta) - 1\right) \tag{S.8}$$

And thus we have:

$$a(\theta) = \frac{\lambda_T p(\theta) + 2F_0 \frac{\pi\lambda_H}{n_S}}{2F_0\left(\lambda_E p(\theta) + \frac{\pi\lambda_H}{n_S}\right)}. \tag{S.9}$$

This is the expression for the modulation. Notice that if the homeostasis weight dominates the others, then the modulation tends to 1. This implies that the adapted tuning curve is identical to the reference tuning curve. If no homeostasis is instead present ($\lambda_H = 0$), the modulation is a constant whose value is fixed by a trade-off between energy and detection weights. In a regime in which $\lambda_E \gg \lambda_H \gg \lambda_T$, the law is a power with exponent -1. We can rewrite the modulation in a more compact form:

$$a(\theta) = \frac{\frac{\lambda_T}{2F_0} p(\theta) + p_U \pi\lambda_H}{\lambda_E p(\theta) + p_U \pi\lambda_H} = \frac{1 + \frac{\lambda_T}{2F_0\pi\lambda_H}\frac{p(\theta)}{p_U}}{1 + \frac{\lambda_E}{\pi\lambda_H}\frac{p(\theta)}{p_U}} \tag{S.10}$$

in which we have factored out $2F_0$ and substituted $p_U = 1/n_S$. The average population response is proportional to the modulation, under the hypothesis of stimulus dependent modulation of equation (S.3):

$$r(\theta) := \frac{1}{n_R}\sum_j r(\varphi_j, \theta) = \frac{1}{n_R}\sum_j a(\theta) r_0(\varphi_j, \theta) = \frac{1}{n_R} a(\theta) \sum_j r_0(\varphi_j, \theta) \propto a(\theta). \tag{S.11}$$

Equation (S.11) implies that if the modulation is a power of the stimulus probability, then the rate is a power of the probability too, $p(\theta)^\beta \propto a(\theta) \propto r(\theta)$. In the next section we show that the modulation is invariant under environment changes and approximately equal to a power law for an extended region of the parameters.

The relation between average population responses and probabilities is invariant under environment change.

Consider equation (S.10). The modulation $a(\theta)$ does not directly depend on the stimulus value, but only on its probability. We can thus rewrite the modulation as a function of the probability of stimuli:

$$a(p) = \frac{1 + \frac{\lambda_T}{2F_0\pi\lambda_H}\frac{p}{p_U}}{1 + \frac{\lambda_E}{\pi\lambda_H}\frac{p}{p_U}}. \tag{S.12}$$

The functional form of $p(\theta)$, thus, is irrelevant to determine the modulation. This implies that the modulation is invariant to environment probability changes. The average population response is also invariant to environment probability changes, since it is proportional to the modulation, see eq. (S.11). This proportionality holds only under the simplifying assumption that the modulation is mainly stimulus dependent.

Approximate power law for an extended region of the parameters space.

In the reduced model (energy-detection-homeostasis, eq (S.4)), the optimal modulation approximates a power law across a range of stimulus probabilities consistent with experiments and over an extended region of the weight space. To demonstrate this, we studied for which regime equation (S.12) approximates a power law. We reformulated equation (S.12) by defining $\tilde{\lambda}_T := \lambda_T/(2F_0\pi\lambda_H)$ and $\tilde{\lambda}_E = \lambda_E/(\pi\lambda_H)$:

$$a(p) = \frac{1 + \tilde{\lambda}_T \frac{p}{p_U}}{1 + \tilde{\lambda}_E \frac{p}{p_U}}. \tag{S.13}$$

Then, we studied the values of $\tilde{\lambda}_T, \tilde{\lambda}_H$ for which equation (S.13) is approximately a power law. We thus looked for the power law cp^β best approximating $a(p)$, by minimizing the following sum of squares error:

$$\mathcal{L}(\beta, c; \tilde{\lambda}_T, \tilde{\lambda}_H) = \sum_p \left(\log a(p) - \beta \log p - c\right)^2.$$

We numerically evaluated the error $\mathcal{L}^*$ for the optimal power law approximating the modulation law in equation (S.13) in the space $\tilde{\lambda}_T, \tilde{\lambda}_H$. We found from the numerical analyses that the error $\mathcal{L}^*$ for the optimal power law is low along a line constrained by $\log \tilde{\lambda}_E = -\log \tilde{\lambda}_T$ (Supplementary Figure S.2a-b-c), namely for $\tilde{\lambda}_E = 1/\tilde{\lambda}_T$. To gain further insight into this result, we rewrote the logarithm of the modulation $\log a(p)$ as a function of the logarithm of the probability ratio $\log \frac{p}{p_U}$ under the constraint $\tilde{\lambda}_T = 1/\tilde{\lambda}_E$:

$$\log a(p) = \log\left(1 + \frac{1}{\tilde{\lambda}_E}\frac{p}{p_U}\right) - \log\left(1 + \tilde{\lambda}_E \frac{p}{p_U}\right). \tag{S.14}$$

Defining $y = \log \frac{p}{p_U}$ and rewriting the logarithm of the modulation as a a function of y, i.e. $f(y) = \log a(p)$:

$$f(y) = \log\left(1 + \frac{1}{\tilde{\lambda}_E} e^y\right) - \log\left(1 + \tilde{\lambda}_E e^y\right), \tag{S.15}$$

Expanding the modulation around $y = 0$ (i.e., $p = p_U$) we have:

$$f(y) = f(0) + \left(\frac{\frac{1}{\tilde{\lambda}_E}}{1 + \frac{1}{\tilde{\lambda}_E}} - \frac{\tilde{\lambda}_E}{1 + \tilde{\lambda}_E} \right) y + O(y^2). \tag{S.16}$$

The expansion variable $y = \log \frac{p}{p_U}$ quantifies the degree with which the probability distribution p deviates from the uniform distribution. If we assume that such deviation is small, we can truncate the Taylor series at its first order on y. Substituting $y = \log \frac{p}{p_U}$ and $f(y) = \log a(p)$, we obtain:

$$\log a(p) = \log a(p_U) + \left(\frac{1 - \tilde{\lambda}_E}{1 + \tilde{\lambda}_E} \right) \log \frac{p}{p_U}, \tag{S.17}$$

where we have also simplified the difference of the two fractions in the expansion. Evaluating the exponential of both sides of the equation we get:

$$\frac{a(p)}{a(p_U)} = \frac{r(p)}{r_U} = \left(\frac{p}{p_U} \right)^{\beta_T(\tilde{\lambda}_E)}, \tag{S.18}$$

where we have used that $a(\theta) \propto r(\theta)$ and where:

$$\beta_T(\tilde{\lambda}_E) = \frac{1 - \tilde{\lambda}_E}{1 + \tilde{\lambda}_E}, \tag{S.19}$$

with the subscript T indicating that the result follows from the Taylor expansion. We compared the result for the exponent of the Taylor expansion with the exponent of the best power law fitting the modulation and we found almost perfect agreement (Supplementary Figure S.2e).

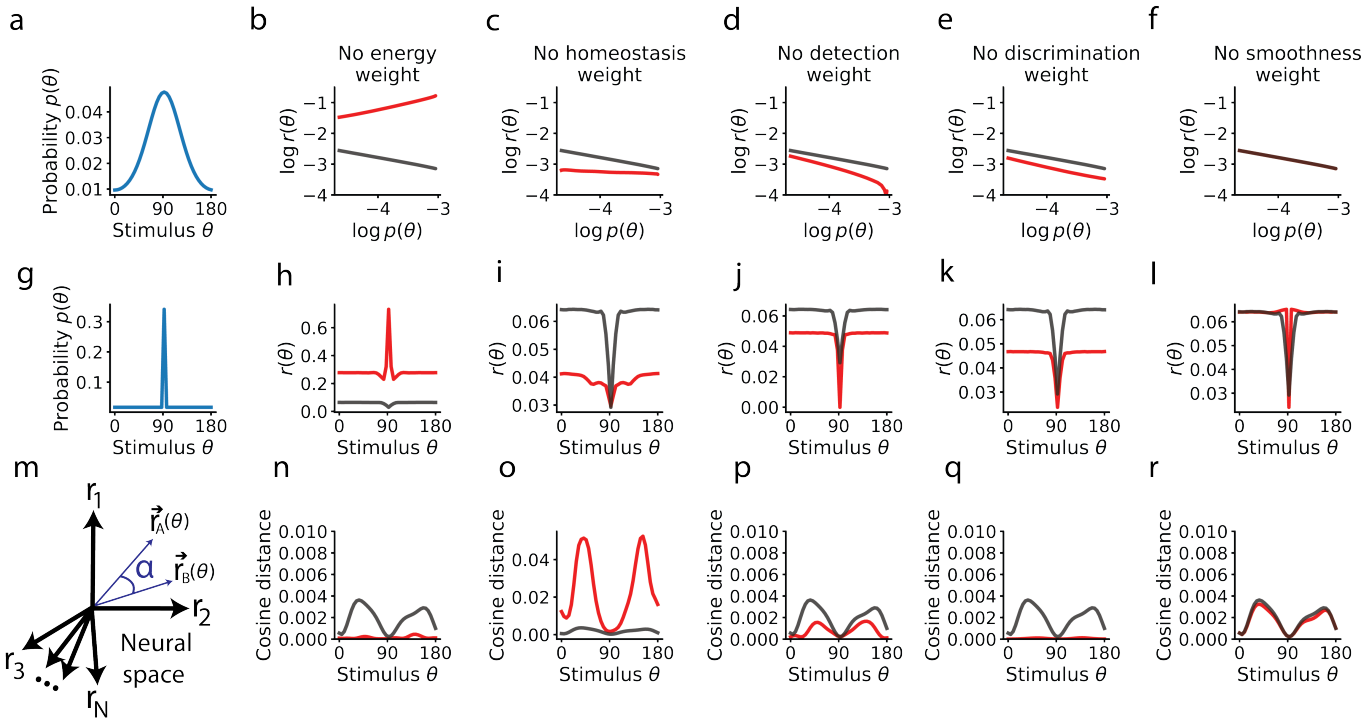

Figure S.1: **All five model costs are necessary to fully replicate the experimental findings.** Only three of the five costs, energy, detection, and homeostasis, are necessary to reproduce a power law. However, all five costs are required to account for the full set of experimental findings. First row: a) Probability distribution of stimulus orientation defining a smooth sensory environment. b-f) Adaptation to environment in panel (a) predicted by the full model (gray lines), or by the modified model with one weight set to zero at a time (red lines): energy (b), homeostasis (c), detection (d), discrimination (e) and smoothness (f). g) Peaked sensory environment. h-l) Response rate as function of the angular stimulus predicted by the full model (gray lines) or the model with one weight set to zero (red lines) to the peaked distribution in (g). Weights are removed in the same order as in (b-f). m) A sketch in high dimensions of the experimental finding, in [1], about the presence of an angle α between adapted response vectors $\mathbf{r}_A(\theta)$ and $\mathbf{r}_B(\theta)$ to the same stimulus θ but in two different environments A and B. n-r) Cosine distance of the angle between two adapted response vectors as a function of the stimulus θ, as predicted by the full model (gray lines) or the model with one weight set to zero (red lines). We computed the cosine distance between two response vectors, adapted to two different environments A and B, responding to the same stimulus θ: $\mathrm{d}_{A,B}(\theta) = 1 - \frac{\mathbf{r}_A(\theta)\cdot\mathbf{r}_B(\theta)}{\|\mathbf{r}_A(\theta)\|\|\mathbf{r}_B(\theta)\|}$. Weights are removed in the same order as in (b-f).

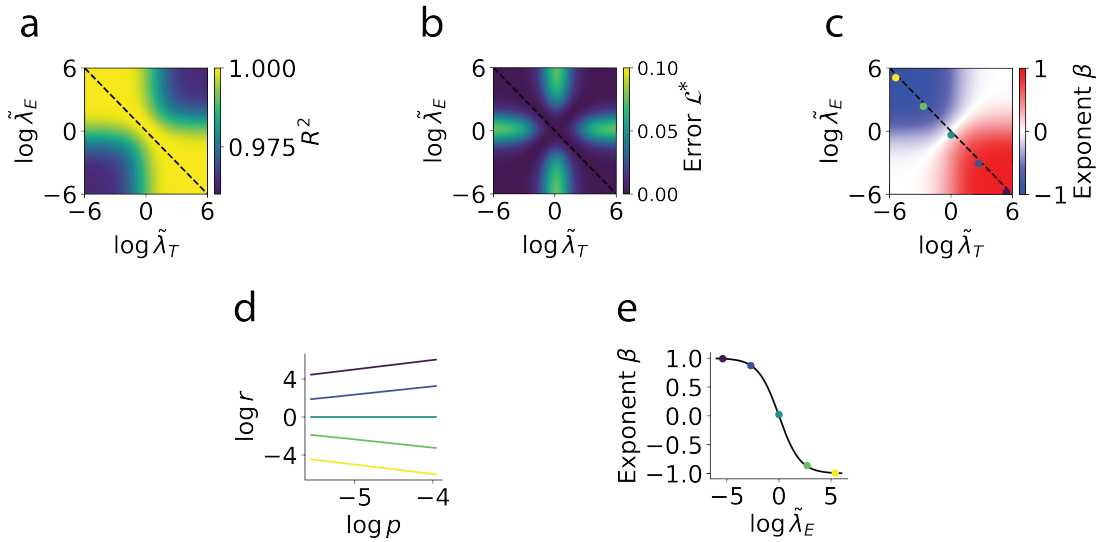

Figure S.2: **An analytic approach finds an approximate power law for the response rate.** a) The R^2 score of the power law approximation of the modulation, shown as a function of the detection and energy weights rescaled by the homeostasis weights: $\tilde{\lambda}_T := \lambda_T/(2F_0\pi\lambda_H)$ and $\tilde{\lambda}_E = \lambda_E/(\pi\lambda_H)$. *Dashed*: the line of parameters which give the best approximation for the power law (maximal R^2 and minimal $\mathcal{L}^*$). It is the line subject to the constraint $-\log\tilde{\lambda}_E = \log\tilde{\lambda}_T$. b) Similar to (a), but for the $\mathcal{L}^*$ error, which is the minimum mean squared error in approximating the rate with a power law. *Dashed:* same of (a). c) The exponent of the best power law fitting the modulation. *Dashed*: same of (a). Colored dots: points on the line of best approximation of power law. d) Several relations between probability and average population response, each one corresponding to a different color in the previous panel. e) Colored dots: the exponent corresponding to each point marked in panel (c). Solid line: exponent predicted by the Taylor expansion of the average population response around a uniform probability.

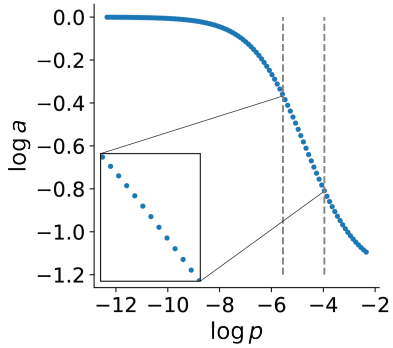

Figure S.3: **The model predicts approximately a power law for values of the stimulus probabilities within the experimental range**. Dotted blue line: the logarithm of the modulation given by equation (S.12) as a function of the logarithm of the probability. Dashed gray lines: interval of probabilities studied in the experiments of [1]. *Inset:* the portion of the modulation law in the probability range studied in the experiment.

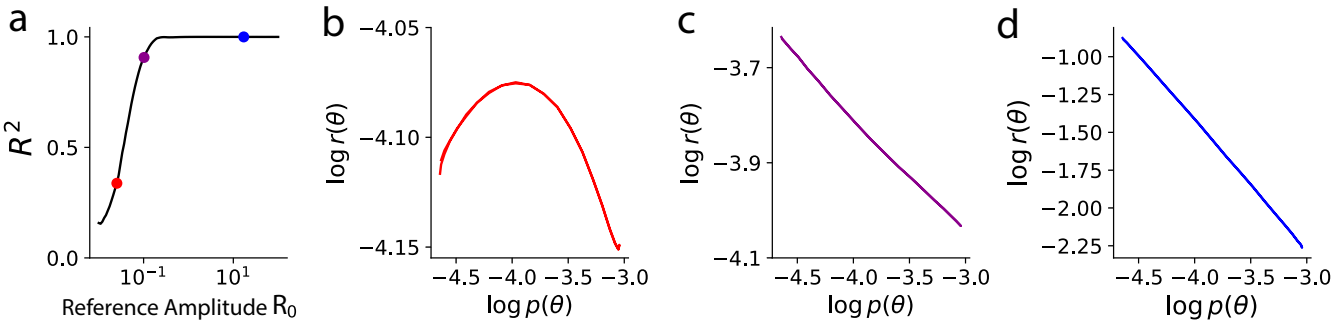

Figure S.4: **The reference tuning curve amplitude R_0 must be sufficiently large for a power law to emerge**. a) The quality of the power law fit, measured with the R^2 estimator, as a function of the initial amplitude r_0 of the tuning curve. (b-d) the representation of the law corresponding to each colored point in panel (a).

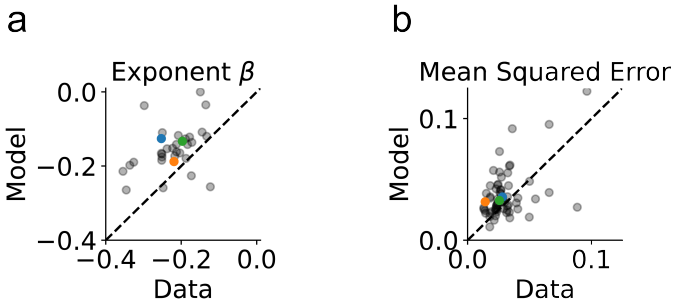

Figure S.5: **The model is robust in predicting the power law exponent for real data.** a) The power law exponent predicted by the model vs. the power law exponent computed from the data (black transparent dots). Colored dots: each color correspond to one of the three instances presented in Figure 4, respectively: blue for (i), orange for (ii) and green for (iii). b) The mean squared error in approximating the response rate to a new environment using the model's predictions and the power law computed from the data. The colored dots have the same meaning of panel (a).

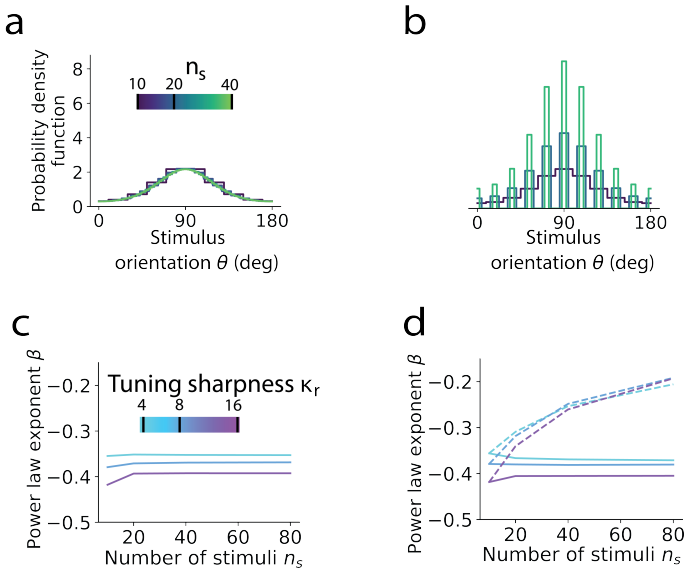

Figure S.6: **Discretization of a continuous environment is different than an environment of discrete stimuli.** a-b) Probability density function for several values of total number of discretization steps n_S in the stimulus space for a continuous (a) and discrete (b) environment of stimuli. In (a), the density converges to a continuous function. In (b), the density converges to a sum of Dirac's deltas, weighted by the probability of each stimulus. c) The power law exponent predicted by our model for the continuous environment in panel (a), as a function of the number of discretization steps n_S in the stimulus space and for several values of tuning curve sharpness κ_r. The exponent is approximately constant for large values of n_S. d) The power law's exponent predicted by our model for the discrete environment of stimuli in panel (b) strongly depends on n_S (dashed lines), showing that our model's weights must be correctly rescaled for a discrete environment. Rescaling the homeostasis weight, $\lambda_H \to \lambda_H \cdot n_S$, leads to the power law's exponent being constant for large n_S (solid lines).