PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Encoding surprise by retinal ganglion cells

Danica Despotović, Corentin Joffrois, Olivier Marre, Matthew Chalk*

Institut de la Vision, INSERM, CNRS, Sorbonne Université, Paris, France

* matthew.chalk@inserm.fr

Abstract

The efficient coding hypothesis posits that early sensory neurons transmit maximal information about sensory stimuli, given internal constraints. A central prediction of this theory is that neurons should preferentially encode stimuli that are most surprising. Previous studies suggest this may be the case in early visual areas, where many neurons respond strongly to rare or surprising stimuli. For example, previous research showed that when presented with a rhythmic sequence of full-field flashes, many retinal ganglion cells (RGCs) respond strongly at the instance the flash sequence stops, and when another flash would be expected. This phenomenon is called the 'omitted stimulus response'. However, it is not known whether the responses of these cells varies in a graded way depending on the level of stimulus surprise. To investigate this, we presented retinal neurons with extended sequences of stochastic flashes. With this stimulus, the surprise associated with a particular flash/silence, could be quantified analytically, and varied in a graded manner depending on the previous sequences of flashes and silences. Interestingly, we found that RGC responses could be well explained by a simple normative model, which described how they optimally combined their prior expectations and recent stimulus history, so as to encode surprise. Further, much of the diversity in RGC responses could be explained by the model, due to the different prior expectations that different neurons had about the stimulus statistics. These results suggest that even as early as the retina many cells encode surprise, relative to their own, internally generated expectations.

Author summary

A long-standing theory suggests that sensory neurons work to transmit as much visual information as possible using little energy. According to this idea, sensory neurons should put most energy into encoding stimuli that are unexpected or surprising. Previous research has shown that this may be true for certain neurons in the retina, that respond strongly when they see something unusual, like a sudden change in a sequence of flashes. Here, to test this theory, we presented the retina with random sequences of light flashes, and measured whether retinal neuron responses correlated with how surprising each flash in the sequence was. Consistent with the theory, we found that retinal neuron responses in our experiment could be explained by a model that assumed that they encode surprise. Moreover, differences between each neuron's responses could be predicted due to their different expectations about which stimuli were most likely.

OPEN ACCESS

Citation: Despotović D, Joffrois C, Marre O, Chalk M (2024) Encoding surprise by retinal ganglion cells. PLoS Comput Biol 20(4): e1011965. https://doi.org/10.1371/journal.pcbi.1011965

Editor: Christoph Mathys, Scuola Internazionale Superiore di Studi Avanzati, ITALY

Received: February 17, 2023

Accepted: March 3, 2024

Published: April 17, 2024

Data Availability Statement: The data is available on: https://github.com/retinal-information-processing-lab/encoding_surprise.

Funding: This work was funded by Agence Nationale de la Recherche (ANR) (grant number OPC-17-CE37-0013-01 and RETNET4EC to MC); OM was supported by ERC Consolidator grant DEEPRETINA (101045253), ANR grants (ANR-18-CE37-0011-DECORE, ANR-20-CE37-0018-04-Shooting star, Chaire Industrielle MyopiaMaster, project NUTRIACT, project PerBaCo, RetNet4EC), a grant from AVIESAN-UNADEV, and from Retina France. This work was supported by the

Programme Investissements d'Avenir IHU FOReSIGHT 497 (ARN-18-IAHU-01). The funders played no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. DD received salary from the ANR (grant OPC-17-CE37-0013-01).

Competing interests: The authors have declared that no competing interests exist.

Introduction

Visual scenes are highly correlated, both in space and time. It has been hypothesized that neurons in early sensory areas have evolved to exploit this structure, by only encoding 'surprising' sensory signals, that cannot be predicted based on their spatio-temporal context. This efficient coding theory can account for many qualitative aspects of neural responses in early sensory areas, such as the stimulus selectivity of neurons in the retina [1–3], as well as primary visual [4–6] and auditory [7, 8] cortices.

A central prediction of the efficient coding theory is that neurons should best encode stimuli that are surprising, given the recent stimulus history. There appears to be some evidence for this in early visual and auditory areas, where neurons have been found that respond most strongly to rare or surprising stimuli [9, 10]. In the retina, previous studies found that when a sequence of full-field light flashes are presented, many neurons respond most strongly at the moment the sequence of flashes is stopped, and where another flash would be expected. This phenomenon was labelled the 'omitted stimulus response' (OSR), since it can be considered to be a response to the stimulus that was unexpectedly omitted [11].

However, if neurons in the retina really do encode surprise, then their responses should vary in a graded way as one varies the level of stimulus surprise. Unfortunately previous studies [11–13] typically measured neural responses to a limited range of flash sequences (e.g. repeated sequences of n consecutive flashes presented in a row). As a result, it is hard to conclude from these studies whether neurons in the retina encode surprise.

To address this question, we presented retinal ganglion cells (RGCs) with extended sequences of stochastically occurring full-field flashes. This extended stimulus included many different sequences of flashes and silences, while the degree of 'surprise' for each flash (or period of silence between flashes) could be quantified mathematically, depending on the previous sequence of flashes and silences. We could thus test how RGC responses varied with the level of surprise. Interestingly, we found that the responses of RGCs to these stimulus sequences could be well explained by a simple normative model, which described how neurons optimally combined their prior expectations about the stimulus with the recent stimulus history to encode surprise. Further, we found that much of the diversity in the responses of different recorded RGCs could be explained by this model, due to the different levels of 'confidence' that different neurons had in their prior expectations. Future work will be needed to elucidate the underlying neural mechanisms that underly these computations.

Our study provides support for the predictive coding model of retinal coding, while shedding light on the different prior expectations that different RGCs have about the environment. More generally, it shows that, already at the stage of the retina, many ganglion cells do not encode the physical stimulus itself, but how unexpected this stimulus is, with different prior expectations for different cells.

Materials and methods

Experimental setup

The recordings were performed in the axolotl retina, using a multi-electrode array with 252 electrodes with 60 μm spacing (procedure described in detail in [14]). The experiment was performed in accordance with institutional animal care standards of Sorbonne Université. The raw signal, recorded at 20 kHz sampling rate, was high-pass filtered at 100 Hz and then sorted offline using SpyKing Circus software [15]. The stimulus consisted of full-field dark flashes. The reason for using dark flashes was the dominance of OFF type cells in axolotl. The dark

flashes had a duration of 40 ms, with 80 ms period between the flashes, (∼ 12 Hz frequency) (as in [11]).

Stimulus statistics

We generated sequences of flashes and silences (i.e. where no flash occured in a 120ms window) using a stochastic model. The number of flashes and periods of silent states presented in a row was drawn from a negative binomial distribution, with parameters r and p. In the case of flashes, we varied the first parameter, p, at 20 minute intervals between three different values (0.98, 0.8 and 0.01, consecutively). The second parameter, r was adjusted so as to maintain a constant mean, of 7 flashes presented in a row. The length of the silence sequence was drawn from a geometrical distribution with a fixed mean p ($p = 9$). Changing p alters the degree to which the distribution is clustered around the mean. However, we observed no difference in the neural responses recorded with different values of p. As a result we concatenated data from neural responses to all three stimulus distributions for the rest of our analysis.

Data analysis

To generate the spike raster plots shown in Fig 1, we aligned the spiking responses of neurons to a sequence of n flashes presented in a row. The peri-stimulus-time-histogram (PSTH) plotted in the bottom row of Fig 1 was computed by averaging the spike count recording over all the stimulus repeats, and then averaging over a 5 ms time bin.

For the remainder of the analysis, we discretised the neural responses and stimulus into time bins of length 120 ms (the time between consecutive flashes). The stimulus presented in each time-bin was treated as a binary variable: '1' if there was a (dark) flash, '0' otherwise. The average firing rate in each bin was computed by average the spike count over all repetitions of a stimulus sequence of length n. Except where stated explicitly in the text, we set $n = 8$ (so there were 256 distinct stimulus sequences used to compute the average firing rate).

Neural model

For our model, we assumed that at each time-bin, t, neurons fire spikes drawn from a Poisson distribution with mean, λ_t, given by:

$$\lambda_t = f(as_t + b) \tag{1}$$

where s_t is the encoded surprise at time t, f is a non-linearity, and a and b are parameters describing the gain and bias, respectively. The non-linearity, $f(x) = \log(1 + e^x)$ (soft-ReLU), was kept fixed for all the cells.

The surprise at time t is defined as:

$$s_t = -\log p(x_t|x_{t-1}, x_{t-2}, \ldots, \theta) \tag{2}$$

where $p(x_t|x_t, x_{t-2}, \ldots, \theta)$ is the probability of observing no flash or a flash at time t ($x_t = 0$ or 1 respectively) given the stimulus x at previous times, and the internal model of the cell, parameterized by θ.

Internal model

The computed surprise depends on each cell's internal model of the stimulus statistics. We first considered a binary Markov model, where the probability of observing a flash at time t is assumed to depend only on whether a flash was observed in the previous time bin. This model has two parameters: $\theta_0 = p(x_t = 1|x_{t-1} = 0)$, and $\theta_1 = p(x_t = 1|x_{t-1} = 1)$. For the Markov 2 model,

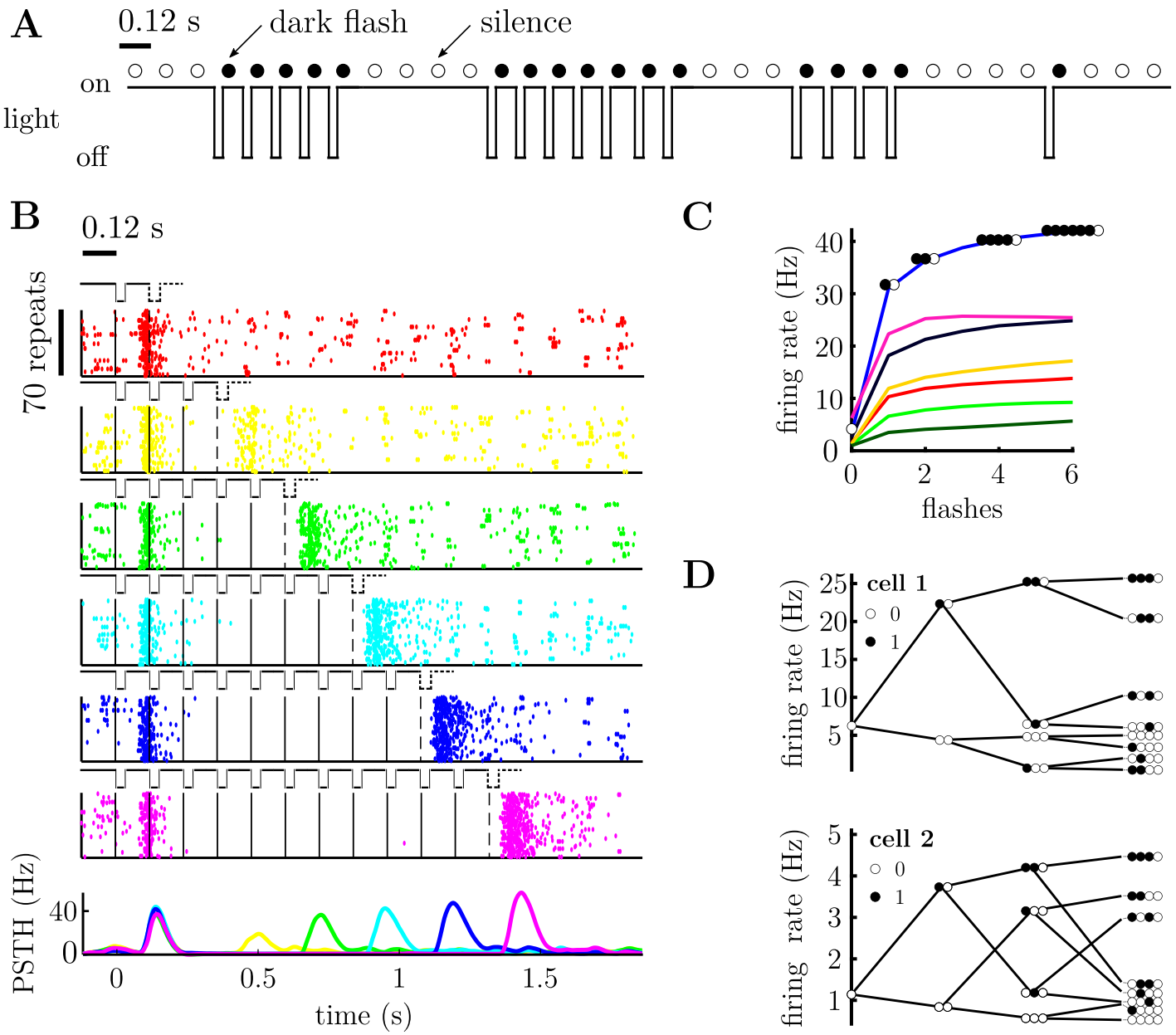

Fig 1. Retinal ganglion cells' responses to a sequences of dark flashes and silences. A. Stimulus excerpt, showing periodic sequences of dark flashes. Each flash lasts 40 ms with 80 ms between. The 120 ms bin containing a single dark flash is marked with a filled circle. A bin without a flash (called a 'silence') is marked by an open circle. **B.** Raster plot for one cell. A solid vertical line marks the occurrence of each flash; a dashed line indicates an 'omitted' flash, following a sequence of flashes. Each raster plot shows the cell's response to a different number of consecutive flashes (ranging from 1 to 11), with 70 repeats shown in each row of the raster. The bottom row shows the peri-stimulus time histogram (PSTH) for different numbers of consecutive flashes (colors denote the number of flashes). There is an increase in firing rate after the missing flash, called the omitted stimulus response (OSR). The OSR magnitude increases with the number of flashes. **C.** OSR for 7 cells, following a varying number of consecutive flashes (filled circles) followed by silence (open circles). **D.** Tree-plot, showing the mean response of two representative cells to different sequences of flashes (filled circles) and silences (empty circles). Cell 1 is the cell plotted in pink in panel C. Each column of the tree-plot shows the average response of the neuron to all stimulus sequences of a given length that end with silence (tree-plots corresponding to sequences ending with a flash are shown in S1 Fig). Moving right-ward the tree-plot branches out to include the effect of stimuli presented further in the past. The top branch of the tree-plot shows the cells' responses to a series of consecutive flashes followed by silence, as in panel C. Other branches show the cells' responses to all the different possible flash sequences of a given length.

https://doi.org/10.1371/journal.pcbi.1011965.g001

we simply extend the observed history to 2 previous states, yielding a total of 4 parameters: $\theta_0 = p(x_t = 1|x_{t-1} = 0, x_{t-2} = 0)$, $\theta_1 = p(x_t = 1|x_{t-1} = 1, x_{t-2} = 0)$, $\theta_2 = p(x_t = 1|x_{t-1} = 0, x_{t-2} = 1)$, and $\theta_3 = p(x_t = 1|x_{t-1} = 1, x_{t-2} = 1)$.

Inferring the transition probabilities

Next, we considered an 'adaptive belief model' where the transition probabilities, $\theta_i = p(x_t|x_{t-1} = i)$, are not known in advance, but must be inferred by combining each cell's prior belief with new observations, using Bayes' law, as follows:

$$p(\theta|x_t, x_{t-1}, \ldots) \propto p(x_t|x_{t-1}, \theta)p(\theta|x_{t-1}, x_{t-2}, \ldots) \tag{3}$$

where $p(x_t|x_{t-1}, \theta)$ is the likelihood of observing x_t given x_{t-1}, and is described by a Bernoulli distribution:

$$p(x_t|x_{t-1} = i, \theta_i) = \theta_i^{x_t}(1 - \theta_i)^{1-x_t}. \tag{4}$$

Let us assume that at time $t - 1$, the posterior distribution over θ_i, $p(\theta|x_{t-1}, x_{t-2}, \ldots)$, is described by the beta distribution with parameters α^i_{t-1} and β^i_{t-1}:

$$p(\theta_i|x_{t-1}, x_{t-2}, \ldots) \propto \theta^{\alpha^i_{t-1}-1}(1 - \theta)^{\beta^i_{t-1}-1}. \tag{5}$$

Now, at time t, multiplying this distribution by the likelihood according to Bayes law (Eq 7) will result in a new beta-distribution, with parameters:

$$\alpha^i_t \leftarrow \alpha^i_t + x_t x^i_{t-1}(1 - x_{t-1})^{1-i} \tag{6}$$

$$\beta^i_t \leftarrow \beta^i_t + (1 - x_t)x^i_{t-1}(1 - x_{t-1})^{1-i} \tag{7}$$

The probability of observing $x_t = 1$ given previous observations is then given by:

$$p(x_t = 1|x_{t-1} = i, x_{t-2}, x_{t-3}, \ldots) = \int_\theta p(x_t = 1|x_{t-1} = i, \theta_i)p(\theta_i|x_{t-1}, x_{t-2}, \ldots) \tag{8}$$

$$= \frac{\alpha^i_{t-1}}{\alpha^i_{t-1} + \beta^i_{t-1}} \tag{9}$$

$$= \frac{n_{i\to 1}}{n_{i\to 1} + n_{i\to 0}}, \tag{10}$$

where $n_{i\to 0}$ and $n_{i\to 1}$ describe the number of occurrences of the transitions $i \to 0$ and $i \to 1$, respectively.

Adaptive surprise model

The statistics of the external world are not static, but change in time. To take this into account, we could assume there is a non-zero probability of transition matrix changing between two observations (a 'dynamic belief model'). Performing exact Bayesian inference in this case requires expensive numerical integration, which may be difficult to perform by individual neurons in the retina. However, in they found that such a dynamic belief model could be approximated by a 'forgetful' model, where recent observations are weighted more strongly than the ones in the past. In contrast to the optimal Bayesian model, their leaky integration model results in simple linear parameter updates, and could thus be easy to implement neurally.

In practice, we can implement the 'forgetful' model of Meyniel et al. [16], by modifying the update rules described earlier for α^i and β^i as follows:

$$\alpha_t^i \leftarrow (1-\eta)\alpha_t^i + \eta\alpha_0^i + x_t x_{t-1}^i (1-x_{t-1})^{1-i} \tag{11}$$

$$\beta_t^i \leftarrow (1-\eta)\beta_t^i + \eta\beta_0^i + (1-x_t) x_{t-1}^i (1-x_{t-1})^{1-i} \tag{12}$$

where $1 > \eta > 0$ is a leak term that results in forgetting observations far in the past, while α_0^i and β_0^i determine the steady state values of α_t^i and β_t^i in the absence of new observations. With this update rule, the probability of observing a $x_t = 1$ given $x_{t-1} = i$ is given by:

$$p(x_t = 1 | x_{t-1} = i, x_{t-2}, x_{t-3}, \ldots) = \frac{\tilde{n}_{i\to 1} + \alpha_0^i}{\tilde{n}_{i\to 1} + \tilde{n}_{i\to 0} + \beta_0^i}, \tag{13}$$

where $\tilde{n}_{i\to j}$ is the 'effective' number of observations of a transition $i \to j$, after taking into account the leak, when $\eta > 0$:

$$\tilde{n}_{i\to j} = \sum_{k=0}^{\infty} (1-\eta)^k x_{t-k}^j x_{t-1-k}^i (1-x_{t-k})^{(1-j)} (1-x_{t-1-k})^{(1-i)}. \tag{14}$$

We assumed that the leak, η was the same for all cells. We used a value of $\eta = 0.2$. However, similar results were obtained when we increased or decreased the leak by a small amount.

In contrast, the parameters of the prior, α_0^i and β_0^i, were allowed to vary for different cells. This allowed us allowed to investigate how different cells' 'prior expectations' for different transitions affected their responses. (Note that for notational simplicity we dropped the subscript '0' in the main text.).

In S1 Text materials we describe a numerical implementation of the exact Bayesian inference model, which assumes that there is a constant probability that the transition probabilities of the Markov model change at each time-step. As mentioned, the results of this model were very similar to our approximate model, described above (S5 Fig).

Model fitting

We fitted the internal model parameters (see previous section), and the gain and bias of the response curves (Eq 4) using Maximum Likelihood (ML) algorithm [17]. For this, we assumed a Poisson noise model, resulting in a log-likelihood:

$$\mathcal{L} = \sum_t n_t \log f_t - f_t \tag{15}$$

where f_t and n_t are the spike count predicted by the model and observed spike count at time t, respectively. All models were fitted using algorithms with multiple starting points (MultiStart in MATLAB, 50 starting points, random initial parameters).

The data analysis and model fitting were done in MATLAB R2021a. Code and data will be available upon paper acceptance.

Results

RGC responses to flash sequences

We used a multi-electrode array to record retinal ganglion cells (RGCs) of an axolotl. We presented a visual stimulus, consisting of random sequences of full-field dark flashes, interleaved

with periods of silence (Fig 1A; see Methods: Stimulus statistics for details). Recorded neural activity was sorted into single unit responses using SpyKing Circus [15].

We were interested in neurons that exhibited an 'omitted stimulus response' (OSR), where they responded to the absence of a flash, following several flashes presented in a row [11]. We thus selected 48 out of 114 single unit responses for further analysis, that showed (i) high quality recording (quantified by low number (<1%) of refractory period violations, where refractory period is 2 ms), and (ii) the presence of an OSR (quantified as a peak around 120 ms after the omitted flash).

Fig 1B shows the example responses of one of these cells to a varying number of flashes presented in a row. As can be seen, this cell responded strongly to the first flash in a sequence, and shortly after the sequence had ended (i.e. the OSR). The size of the OSR increased monotonically with the number of flashes presented in a row.

For our analysis, we converted the stimulus to a binary variable, which was set to 1 or 0 depending on whether there was a dark flash (stim. = 1) or a period of silence (stim. = 0) within a given 120ms window. Neural responses were taken to be the number of spikes that occurred within each 120ms window.

To see how the OSR varied with the number of consecutive flashes, we computed the average response of each neuron, given a 'stimulus history' consisting of a varying number of consecutive flashes followed by silence (Fig 1C). The OSR increased monotonically with the number of flashes for all cells. However, we observed differences in the rate of increase as well as the maximum firing rate for different cells (Fig 1C).

Finally, to see how neural responses depended on all possible stimulus sequences (and not the number of consecutive flashes), we constructed 'tree-plots' (Fig 1D), showing each neuron's average response to all possible stimulus sequences of a given length, ending with a silence (the complementary tree-plot, for sequences ending with a flash, is shown in S1 Fig). The top branch of this tree plot corresponds to the OSR, shown in Fig 1D. However, many cells that showed a qualitatively similar OSR (i.e. that increased with the number of consecutive flashes) exhibited very different tree-plots, identifying clear differences in how they responded to different patterns of flashes and silences (e.g. Fig 1D).

Modeling 'surprise encoding' by RGCs

We asked whether RGC responses were consistent with them encoding surprise. To test this, we constructed a simple model of how RGCs could combine their internal stimulus expectations with their recent stimulus history to compute surprise (Fig 2A). Following [18, 19], we defined surprise at time t, s_t, as the negative log probability of a stimulus, x_t, given the recent stimulus history, $x_{<t}$, and the neuron's internal model of the stimulus statistics (parameterised by θ):

$$s_t = -\log p(x_t|x_{<t}, \theta) \tag{16}$$

The mean firing rate was then obtained by applying a simple non-linear mapping:

$$r_t = f(as_t + b) \tag{17}$$

where a and b are free parameters and $f(\cdot)$ was assumed to be a softplus ($\log(1 + e^x)$) non-linear function to prevent firing rates being negative.

The computed 'surprise' for each cell thus depends on their expectations or 'internal model' of the stimulus statistics (parameterized by θ). We first assumed the simplest possible internal model: a Markov model, in which the probability of observing a flash, $x_t = 1$, only depends on whether there was a flash or not in the previous time bin ($x_t = 0/1$). We called this the 'fixed

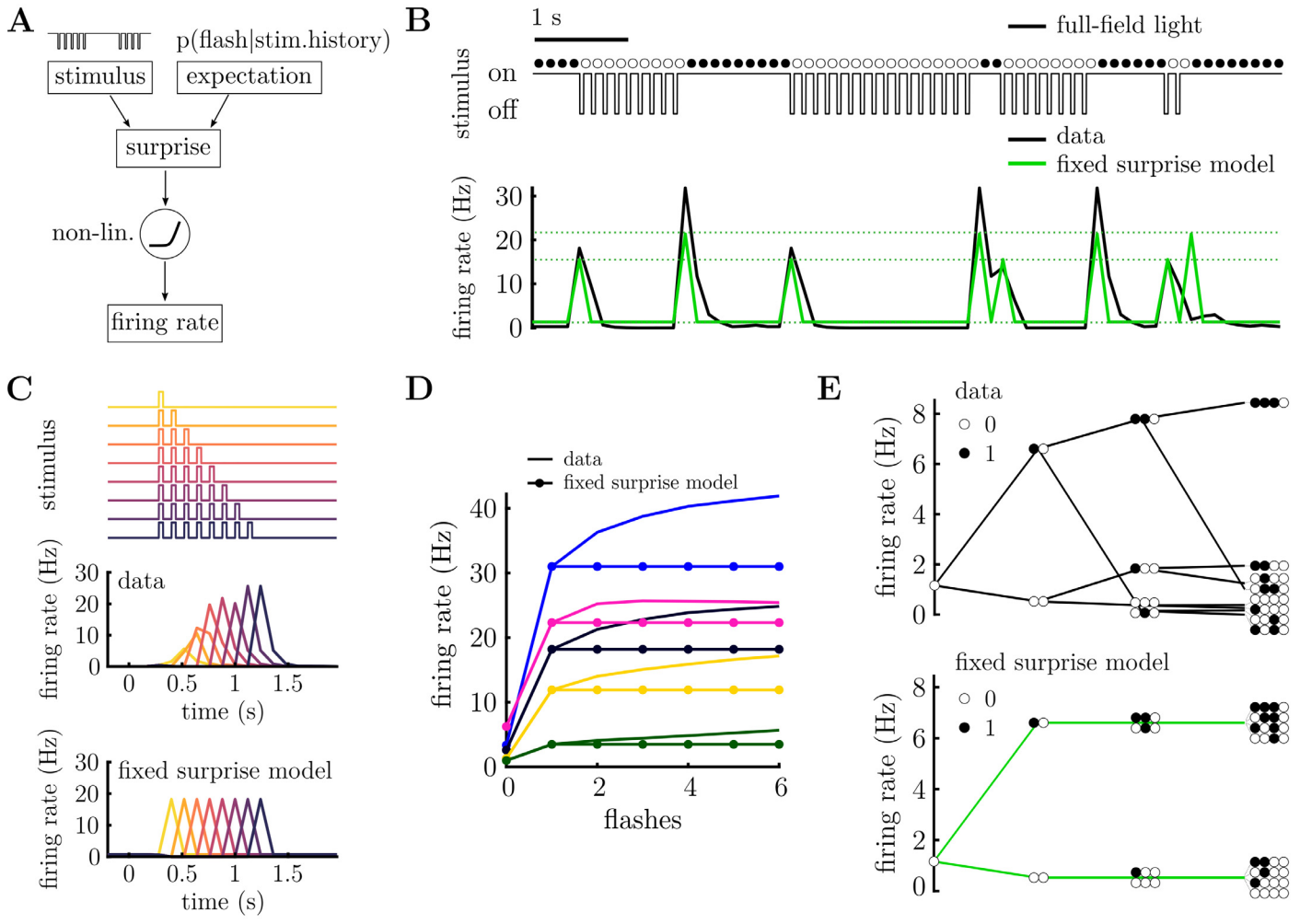

Fig 2. Fixed surprise model. A. Schematic of modeling framework. The stimulus is compared to the neuron's expectation, which depends on their internal model, to compute surprise. The encoded surprise is then transformed via a static non-linearity to obtain the neuron's firing rate. **B.** Stimulus excerpt (above) and recorded PSTH (below, black), and prediction of the fixed surprise model (below, green). The fixed surprise model has limited flexibility, only permitting four possible firing rates (indicated with dashed lines). **C.** Response to flash sequences of varying length (top). PSTH for a single neuron (middle) and model prediction (bottom) to the stimulus sequences shown above. Each colour corresponds to a different length of flash sequence. The fixed surprise model predicts the OSR magnitude to be independent of the number of flashes. **D.** OSR for 5 cells (solid lines) after a varying number of consecutive flashes. The fixed surprise model (lines with filled circles) cannot account for the increase in the OSR with increasing number of flashes. **E.** Tree-plot for a single cell (above) and fixed model prediction (below). The fixed surprise model can capture the mean response for stimulus sequences of up to length 2, but not beyond. Tree-plots corresponding to sequences ending with a flash are shown in S2 Fig.

https://doi.org/10.1371/journal.pcbi.1011965.g002

surprise' model, since the surprise is fixed by the stimulus presented in the previous time-step. This fixed surprise model has two free parameters: the probability of a flash occurring if there was/wasn't a flash in the previous time-step ($\theta_0 = p(x_t = 1|x_{t-1} = 0)$, and $\theta_1 = p(x_t = 1|x_{t-1} = 1)$). The parameters of the response function (a and b) and internal model (θ) were fitted for each neuron using maximum likelihood, assuming that the responses were generated by a Poisson distribution with mean r_t (see Methods: Neural model).

Fig 2B shows the average firing rate of a single neuron (black) for a given stimulus sequence (above) (see Methods: Data analysis). This model accounted for the most prominent feature of the neuron's responses: that it responded strongly to the first flash in a sequence, and the first

silence in a sequence (i.e. the OSR). However, the model was unable to replicate the dependence of the OSR on the number of flashes presented in a row, observed for this (Fig 2C) and many other cells (Fig 2D). This was because, by design, with a Markov model the computed surprise, only depends on the stimulus in the previous time-bin, and thus only two responses to a silence are possible, depending on whether there was a flash or not in the previous time-step (Fig 2E; see S2 Fig for complementary tree-plot, for sequences ending with a flash).

Adaptive surprise model

To account for the observed variations in the OSR with number of consecutive flashes, we next considered a more complex 'dynamic belief' internal model. Here, we assume that the transition probabilities ($\theta_i \equiv p(x_t = 1|x_{t-1} = i)$), are not known *a priori* by each neuron, but must be inferred. We assume neurons combine their prior expectations ($p(\theta_i)$) with the recent stimulus history ($p(x_t, x_{t-1}, \ldots|\theta)$) using Bayes' law: $p(\theta|x_t, x_{t-1}) \propto p(x_t, x_{t-1}, \ldots|\theta_i)p(\theta_i)$. We assumed a beta-distribution for the prior over θ_i, with parameters α_i and β_i of the form: $p(\theta_i) = \theta_i^{\alpha_i - 1}(1 - \theta_i)^{\beta_i - 1}$. Multiplying the beta prior with the Bernoulli distribution, $p(x_t, x_{t-1}, \ldots|\theta_i)$, we obtain:

$$p(\theta_i|x_t, x_{t-1}, \ldots) \propto p(x_t, x_{t-1}, \ldots|\theta_i)p(\theta_i) = \theta^{n_{i\to 1}+\alpha_i-1}(1-\theta)^{n_{i\to 0}+\beta_i-1} \quad (18)$$

where $n_{i\to j}$ is the number of occurrences of the transition $i \to j$ in the sequence $\{x_1, x_2, \ldots, x_t\}$, and α_i and β_i are parameters of the prior. Finally, we integrate out θ, to obtain a simple expression for the inferred probability of observing $x_t = 1$, given $x_{t-1} = i$:

$$p(x_t = 1|x_{t-1} = i, x_{t-2}, \ldots) = \langle\theta\rangle_{p(\theta_i|x_t, x_{t-1}, \ldots)} = \frac{n_{i\to 1} + \alpha_i}{n_{i\to 0} + n_{i\to 1} + \beta_i + \alpha_i}, \quad (19)$$

We assume that the parameters of the prior (α_i, β_i) are different for each neuron. Note that, in the limit where the prior is very strong (i.e. $n_{i\to j} \ll \alpha_i$ and $n_{i\to j} \ll \beta_i$), this model becomes identical to the 'fixed-belief' model described previously (Fig 2), where the inferred transition probabilities do not depend on the stimulus history, beyond the previous time-step.

If neurons had 'infinite' memory then, given a sufficiently long stimulus sequence, their prior expectations would have no effect. Instead, we assume a more biologically plausible model where neuron's have a finite memory, and $n_{j\to i}$ are estimated using a leaky integration of past observations (see Methods: Adaptive surprise model). This requires one additional parameter (the time-scale of integration i.e. the leak parameter), which we kept fixed for all neurons. In S1 Text we show how qualitatively similar results can be obtained by assuming neurons perform Bayesian inference, given a model where the transition probabilities have a small probability of changing on each time-step. However, optimal Bayesian inference in this setting required complex numerical integration, and it is thus hard to see how it could be implemented feasibly by individual neurons. As a result, we focus on the simpler 'leaky integration' model for the rest of the paper.

We fitted the 4 parameters of the prior (plus the gain and bias of the LN model, corresponding to a and b in Eq 2, respectively) for each neuron, using maximum likelihood, assuming Poisson noise [20]. S12 Fig shows the fitted prior for 6 representative cells. Fig 3A shows the predicted firing rate for one neuron (blue) to a short stimulus sequence (above). The 'adaptive surprise' model was able to capture aspects of the neuron's response that could not be accounted for by the previous 'fixed surprise model'. For example, it could capture how the size of the OSR increased with the number of flashes presented in a row (Fig 3B and 3C). Further, it captured individual differences in the OSR decay for different neurons (compare cells 1–3 in Fig 3B). Overall, the correlation between the estimated firing rates and the model

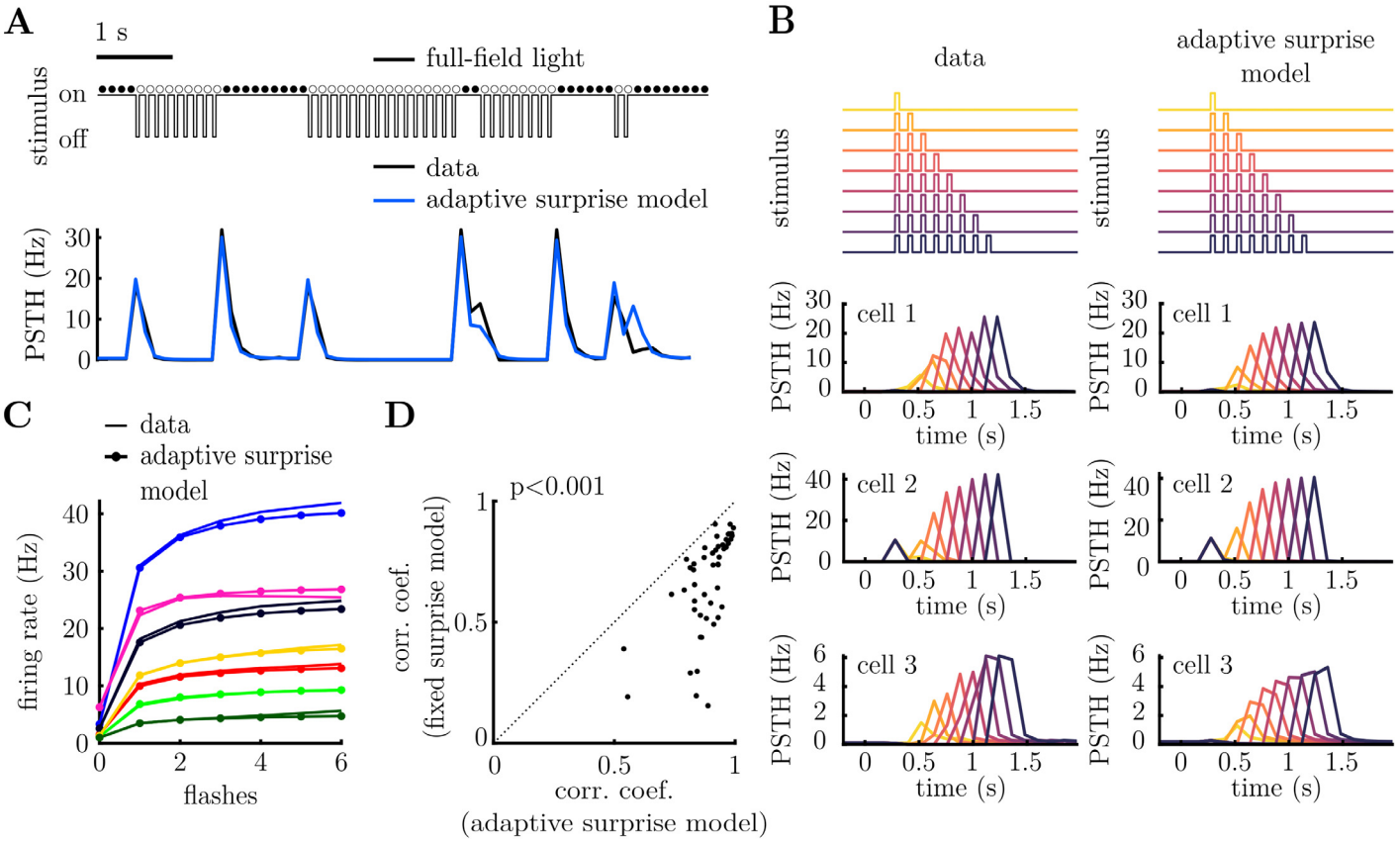

Fig 3. Adaptive surprise model. A. Stimulus excerpt (above) and recorded PSTH (below, black), alongside prediction of the adaptive surprise model (below, blue). **B.** Neural responses to varying number of consecutive flashes (above). Recorded PSTH of three neurons is shown to the left, while model predictions are shown to the right. Each colour corresponds to a different number of consecutive flashes. The adaptive surprise model captures variations in both the magnitude and width of the OSR. **C.** Increase in the OSR with the number of consecutive flashes for seven cells (each cell plotted with a different colour). The data (solid line) is plotted alongside the predictions of the adaptive surprise model (solid lines with circles). **D.** Pearson correlation coefficients between each cell's PSTH and the model predictions, for the fixed surprise model (y-axis) versus the adaptive surprise model (x-axis). The adaptive surprise model significantly outperforms the fixed surprise model ($p = 1 \cdot 10^{-9}$, Wilcoxon signed-rank test).

https://doi.org/10.1371/journal.pcbi.1011965.g003

prediction was significantly higher for the adaptive, compared to the fixed, surprise model (Fig 3D).

To further investigate how the adaptive surprise model could account for the diverse responses of different cells, we plotted tree-plots showing the average firing rates predicted by the model for stimulus sequences of different lengths (Fig 4A; see S3 Fig for complementary tree-plot, for sequences ending with a flash). The adaptive belief model captured much of the structure in the neural responses to stimulus sequences of varying length, as well as the diversity across different cells. This was supported by plotting the correlation coefficient between the model predictions for each node of the tree and the data, which decayed slowly with the tree depth (Fig 4B), compared to the fixed surprise model which reduced dramatically for tree depth greater than 2. Finally, S10 Fig shows that our model can account for how neural responses depend on stimuli further in the past, quantified by how much different nodes branch out from their parent node as we increase the depth of the tree plot.

To further test our adaptive belief model, we compared it to a more complex fixed belief model, with a comparable number of free parameters. To do this, we implemented a 'Markov-2

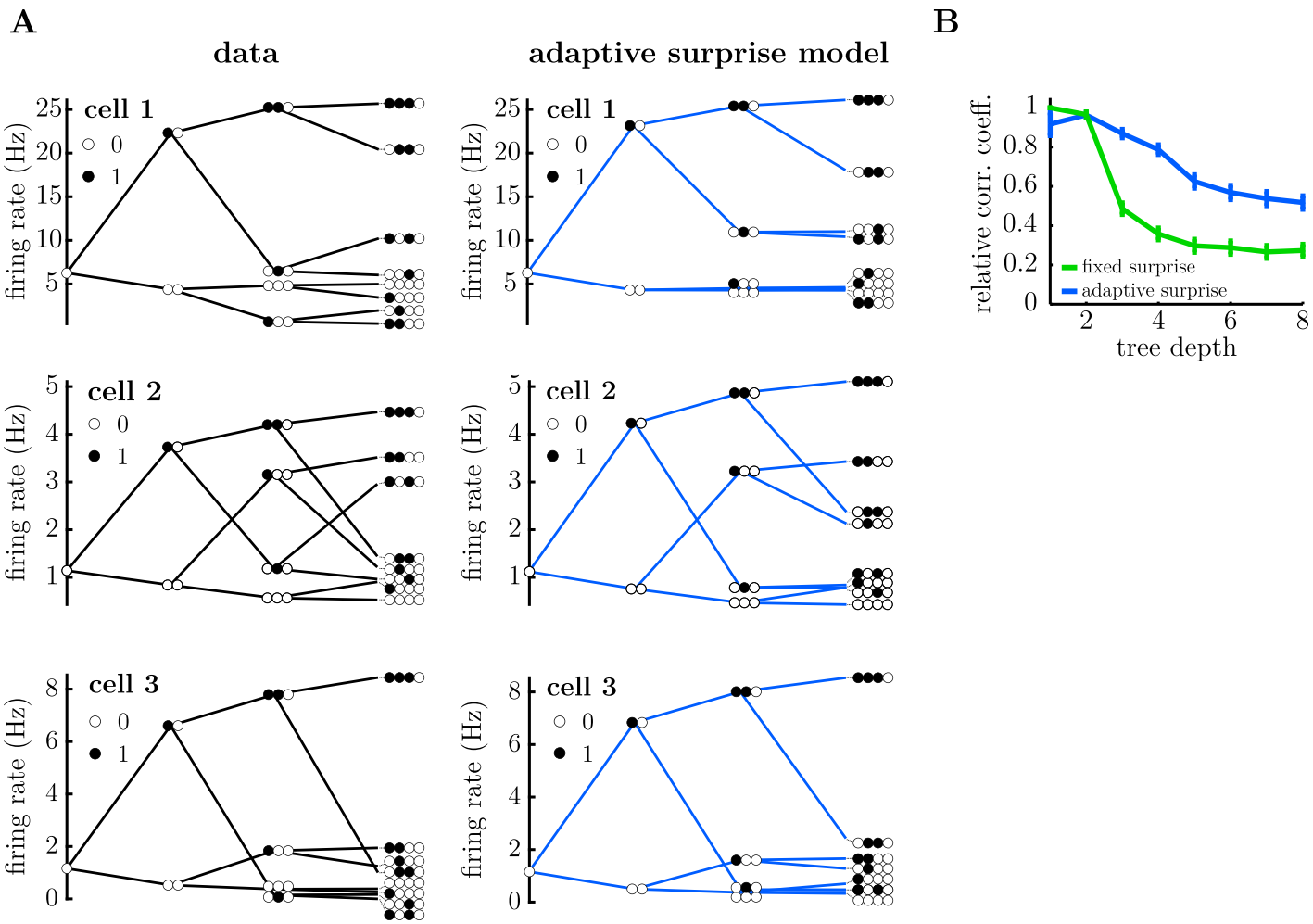

Fig 4. Neural responses to different stimulus sequences, for the adaptive surprise model. A. Tree-plot, showing the mean response of three representative cells to all possible sequences of flashes (filled circles) and silences (empty circles) of a given length. As we move rightward, the tree branches to show responses to take into account stimuli presented further in the past. The data is shown on the left and the model predictions on the right. The adaptive model is a able to reproduce qualitative aspects of each tree-plot, beyond the top branch (which shows how the OSR magnitude varies with the number of consecutive flashes). Tree-plots corresponding to sequences ending with a flash are shown in S3 Fig. **B.** Correlation coefficient between the tree-plot obtained with the adaptive surprise model and the data, computed separately for each tree-depth (i.e. stimulus sequence length). The adaptive model is significantly better at capturing the shape of the tree for stimulus sequences longer than 2, compared to the fixed surprise model.

https://doi.org/10.1371/journal.pcbi.1011965.g004

model' in which the probability of observing a flash is depends on the observed stimulus in the previous two time-bins. This model's prior has 4 parameters, ($\theta_{ij} = p(x_{t+1} = 1|x_t = i, x_{t-1} = j)$), which is the same as the adaptive surprise model (aside from the leak parameter, which we kept the same for all cells). The behaviour of this model is shown in Fig 5. While the Markov-2 model outperformed the fixed surprise model model described earlier, it could not, by construction, account for increases in the OSR that occurred for sequences of more than 2 consecutive flashes (Fig 5B and 5C), or any structure in the tree plots at a depth greater than 2 (Fig 5D, emphasized with a dashed ellipse). Finally, the correlation coefficient between predicted and observed firing rates was significantly worse for the Markov-2 model than the adaptive surprise model (Fig 5E) despite them having the same number of free parameters ($p = 2 \cdot 10^{-8}$, Wilcoxon signed-rank test).

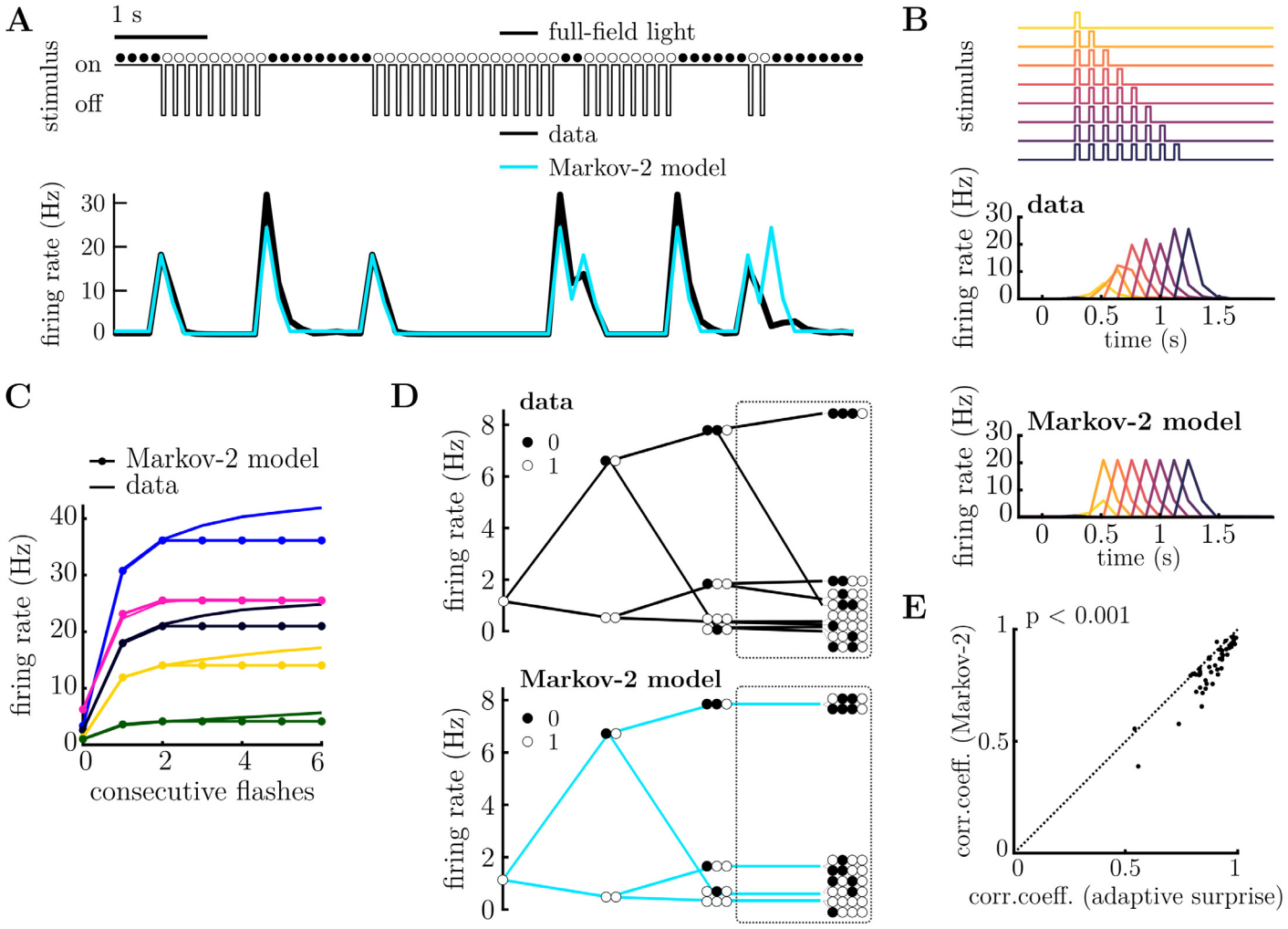

Fig 5. Fixed surprise model with longer past (Markov-2 model). A. Stimulus excerpt (above) and recorded PSTH (below, black), and firing rate predicted by the Markov-2 model (below, blue). **B.** Response to varying number of consecutive flashes (top). PSTH for a single neuron (middle) and model prediction (below) to the stimulus sequences shown above. Each colour corresponds to a different length of flash sequence. The Markov-2 surprise model predicts the OSR magnitude to be dependent on the previous two state only. **C.** Average responses of 5 cells (solid lines) to flash sequences of varying lengths. The Markov-2 surprise model (lines with filled circles) cannot account for the increase in the OSR beyond 2 consecutive flashes. **D.** Tree-plot for a single cell (above) and fixed model prediction (bottom). The Markov-2 surprise model cannot capture the response for stimulus sequences greater than length 2 (highlighted with dashed circle). Tree-plots corresponding to sequences ending with a flash are shown in S4 Fig. **E.** The correlation coefficient between the adaptive surprise model and recorded PSTH for each cell is significantly better than for the Markov-2 model ($p = 2 \cdot 10^{-8}$, Wilcoxon signed-rank test).

https://doi.org/10.1371/journal.pcbi.1011965.g005

Differences in the internal expectations for individual cells

We were interested to see how the inferred expectations (the 'prior') varied for each cell. Recall that we assumed a beta-prior over the transition probabilities $\theta_i = p(x_t = 1|x_{t-1} = i)$, with parameters α_i and β_i. The mean of this prior is determined by the ratio of these two parameters, α_i/β_i, while its width (i.e. the level of prior uncertainty) is determined by their sum, $\alpha_i + \beta_i$. Fig 6A and 6B show how the parameters of the prior varied for different cells. Interestingly, we found that while for different cells there was a large variation in the sum, $\alpha_i + \beta_i$, the ratio, α_i/β_i, was relatively constant. Thus, while the width of the prior, which determines how much weight is accorded to prior expectations versus new observations, varied greatly across cells, the prior mean was roughly constant.

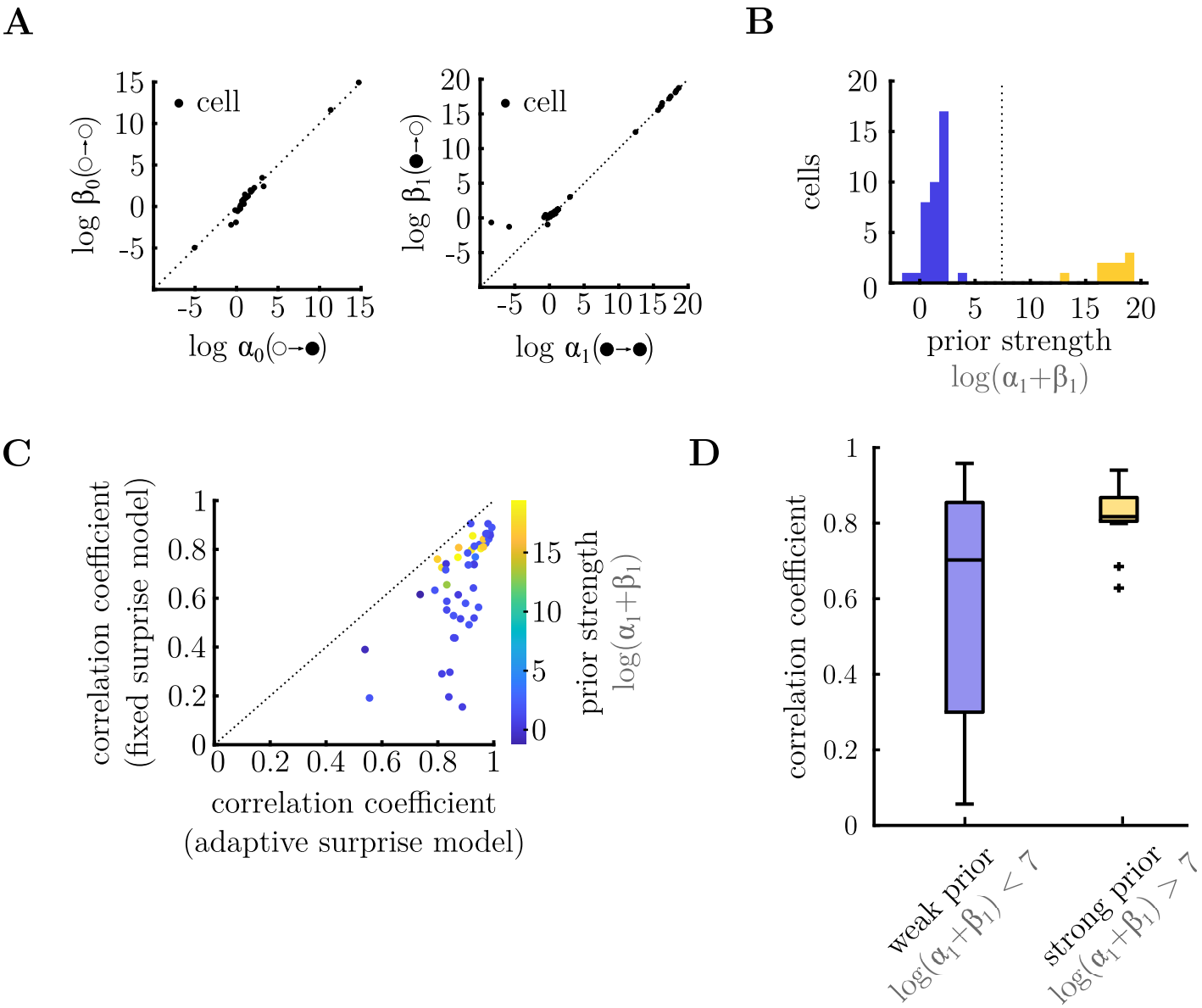

Fig 6. Parameters of internal model. A. Parameters of the inferred prior (α_i and β_i) for each cell. These parameters determine each cell's prior expectation for the different transitions from silence (left) or flash (right). In both cases, the ratio between these parameters, α_i/β_i (which determine the mean of the prior) is close to unity for all the cells, while their sum, $\alpha_i + \beta_i$ (which determines the strength of the prior) varies across different cells. **B.** Histogram of $\log(\alpha_1 + \beta_1)$ for different cells. The population could be split into two groups: cells with low confidence in the prior (i.e. small $\alpha_i + \beta_i$; blue) and cells with high confidence in the prior (i.e. large $\alpha_i + \beta_i$; yellow). **C.** Correlation coefficient between the responses predicted by the adaptive surprise model, versus the fixed surprise model. Each circle is colour coded according to the parameters of the inferred prior for that cell $\log(\alpha_1 + \beta_1)$. Cells that had a strong prior (yellow) tended to be better fit by the fixed surprise model, relative to the adaptive surprise model. **D.** For each cell we computed the correlation coefficient between the average neural responses to stimulus sequences of length 2, versus average responses that take into account stimulus sequences of length 10. A correlation of 1 would imply that neural responses just depended on the stimulus in the previous time-step; values less than 1 imply that neural responses depend on stimuli further in the past. As expected, the responses of cells with a strong prior (right, yellow) exhibited a higher correlation coefficient than cells with a weak prior (left, blue), showing that they could be better predicted just by looking at the most recent stimulus transition.

https://doi.org/10.1371/journal.pcbi.1011965.g006

Focusing on the prior parameters, α_1 and β_1, which determines neural responses to 'flash→flash' and 'flash→no-flash' transitions (i.e. the OSR), we observed two clusters of cells (Fig 6A, right panel), with different levels of prior uncertainty (determined by the sum, $\alpha_1 + \beta_1$; Fig 6B). We asked what effect this would have on these cells responses. We reasoned

that cells with a strong prior (i.e. large $\alpha_i + \beta_i$) would not adapt their posterior belief much depending on recent observations, and hence their responses would be well predicted by the fixed surprise model. In contrast, cells with a weak prior (i.e. small $\alpha_i + \beta_i$) would be strongly influenced by recent observations, and thus their responses would be poorly predicted by the fixed-surprise model. This turned out to be the case. Fig 6C shows the correlation coefficient between the prediction of the fixed surprise model and recorded responses, versus the adaptive surprise model. There was a trend for cells with a strong prior (high $\alpha_i + \beta_i$; colour coded in yellow) to be equally well-fit by both models, while cells with a weak prior (low $\alpha_i + \beta_i$; colour coded in blue) were better fit by the adaptive surprise model.

We next asked whether this effect could be observed in the data, without reference to the model fits. To do this, we compared the average neural responses to stimulus sequences of length 2, to average responses to longer stimulus sequences, of length 10. A correlation of 1 would imply that neural responses just depended on the stimulus in the previous time-step; values less than 1 imply that neural responses depend on stimuli further in the past. As expected, we found that cells with a strong prior (i.e. $\log(\alpha_1 + \beta_1) > 7$) exhibited high correlation coefficients, close to 1, implying that their responses could be well predicted by just looking at the most recently presented stimuli (Fig 6D, yellow). This was not the case for stimuli with a weak prior (Fig 6D, blue) ($p = 0.066$, Wilcoxon rank-sum test).

In Fig 6A we observed that the prior mean, determined by α_i/β_i, remained near-unity across different cells. We thus, asked whether it would be possible to fit neural responses using a reduced adaptive surprise model with only two parameters (i.e. the sum, $\alpha_i + \beta_i$), and α_i/β_i held fixed at unity. Fig 7A shows that, while this reduced adaptive model performed worse than the full adaptive surprise model, this reduction in performance was small (< 10% reduction in correlation coefficient), despite having having only 2 free parameters per cell (compared to 4 parameters, for the full model). Notably, the reduced model was able to capture similar qualitative features of neural responses, such as how the OSR increased with the number of consecutive flashes (Fig 7B and S5 Fig). Its performance was also significantly better than the fixed

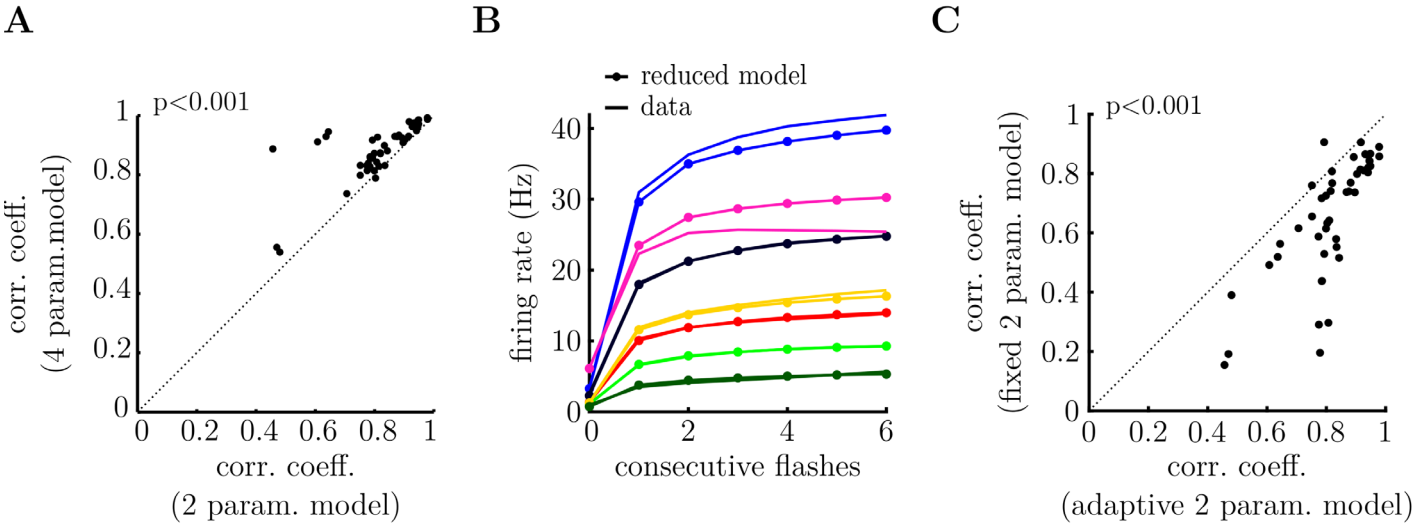

Fig 7. Reduced adaptive surprise model, with a fixed prior mean (i.e. $\alpha_i/\beta_i = 1$). A. The reduced model performs almost as well as the adaptive surprise model despite having half the number of free parameters. ($p = 3 \cdot 10^{-9}$, Wilcoxon signed-rank test). **B.** Mean response of 7 cells following a variable number of flashes presented in a row. The increase in the OSR with the number of flashes is well-fitted by the reduced model. **C.** The reduced adaptive surprise model performs significantly better at fitting the recorded neural responses, despite both models having the same number of free parameters per cell. ($p = 4 \cdot 10^{-5}$, Wilcoxon signed-rank test.).

https://doi.org/10.1371/journal.pcbi.1011965.g007

surprise model, which had the same number of free parameters per cell ($p = 4 \cdot 10^{-5}$, Wilcoxon signed-rank test). Tree-plots for the reduced adaptive surprise model are shown in S4 Fig.

Discussion

We observed how neural responses in the retina showed non-trivial dependencies on the precise order of flashes and silences in random stimulus sequences (Fig 1C and 1D). Interestingly, RGC responses were well predicted by a simple model, which assumed that they depended on how 'surprising' stimuli were, relative to an internally generated expectation (Figs 3 and 4). Moreover, our model showed how the different 'expectations' of different neurons could account for the diverse way they responded to presented stimuli (Figs 6 and 7).

Our approach contrasts with previous ideal observer models, which assume that neurons are perfectly adapted to the 'true' presented stimulus statistics [21–23]. Instead, we found that neural responses could be well explained by assuming that each neuron has learned its own internal model of the stimulus statistics (with the parameters of the prior fitted separately for each cell). Interestingly, we found that different neurons had very similar prior expectations about which stimuli were most likely to occur (determined by the mean of the prior). What varied was the degree of confidence they had about their own prior expectations (determined by the width of the prior). Furthermore, recorded cells could be divided into two categories: those with weak confidence in their prior, and those with strong confidence in their prior expectations. We were interested to see whether this split was related to different functional cell types. To do this, we performed a second experiment, where we presented a full-field 'chirp' stimulus, which has been used previously to classify different cell-types in the mouse [24]. We confirmed that the main results with this second experiment were qualitatively identical (S6 Fig). We then performed clustering based on the responses of different cells to the chirp stimulus (S7 Fig). We did indeed find a trend for some cell types to exhibit an OSR (S8 Fig), as well as for some types to show larger/stronger fitted priors (S9 Fig). However, the fitted prior would be clearly insufficient to identify the cell type alone, since there was a large spread of parameters within each cell type. Further experiments would be required to investigate further the relation between surprise encoding and cell type.

Our modelling framework was adapted from a previous model of Meyniel et al., that sought to explain psychophysical data showing how subjects' behaviour (such as their reaction time and accuracy) depended on the statistics of sequentially presented sensory stimuli [16]. Meyniel and colleagues showed how their data could by explained if subjects used a Bayesian inference model, as described here, to predict new stimuli based on what came before. Here, we extended this model to include a variable 'prior' distribution, whose parameters could be fit to describe the diverse responses of different ganglion cells. Nonetheless, the fact that a similar type of model can be used to describe both neural responses in the retina and subjects behaviour in different tasks is intriguing, raising the question of whether similar computations may be present ubiquitously in the brain when subjects are presented stimuli with complex temporal statistics. This is an important link to explore, since many previous experiments have shown how our perceptual experience is flexible, and can be adapt depending on the recent stimulus history [25, 26].

Previously it was shown that in certain cases, retinal neurons are able to adapt to the statistics of their environment, so as to preferentially encode stimuli that are most unpredictable in any given context [27, 28]. Recently, Mlynarski & Hermundstadt developed a mathematical framework to account for how such adaptive efficient coding could be performed by sensory neurons [29, 30]. The question remains however, over which timescales, and which type of stimulus statistics, retinal neurons are able to adapt to. In our experiment, we found that RGCs

adapt to recent stimulus statistics: notably the sequence of flashes presented over a period of 1–2 seconds. While beyond the scope of the current work, previous studies have proposed several possible mechanisms for this, including depressing inhibitory synapses [31] and combined ON and OFF pathways [13] (S11 Fig). On the other hand, however, we found little evidence that RGCs could adapt to stimulus statistics over longer time-scales. This was evidenced by the fact their learned prior differed significantly from the stimulus statistics presented over the course of the experiment. Further, we found that introducing subtle changes to the stimulus statistics (e.g. increasing the probability of either very long or short flash sequences, while keeping the average number of flashes presented in a row the same), had little effect on the resulting RGC response properties in our experiment. It would be interesting investigate this further in the future to see, for example, how and whether RGC responses vary depending on more strong changes in the stimulus statistics, including in the presence of more complex, naturalistic, stimuli.

Previous experimental [12, 13, 32] and computational [33–35] studies sought to understand the neural mechanisms underlying the OSR. However, there remains some controversy over which of the proposed theories could explain all of the experimentally observed features of the OSR, such as e.g. the fact that the delay before the OSR varies linearly with the time between flashes. Our work provides further constraints to distinguish between different theories, by showing how the OSR varies depending on the precise sequence of flashes and silences (Fig 1).

The stimuli in our experiment, which consisted of sequences of full-field flashes, were chosen to be sufficiently rich so as to permit many different levels of 'surprise', while simple enough to permit a straight-forward analysis of neural responses. Nonetheless, in the future it would be interesting to investigate neural responses to more naturalistic stimuli, which for example, varied spatially as well temporally [36]. This would allow us to investigate, for example, the degree to which neurons' internal model is adapted to the statistics of natural scenes, as predicted by the efficient coding hypothesis [37].

Supporting information

S1 Fig. Tree-plot for two representative cells, corresponding to sequences ending with a flash. Here we show the mean response of two representative cells to different sequences of flashes (filled circles) and silences (empty circles). Each column of the tree-plot shows the average response of the neuron to all stimulus sequences of a given length, that end with flash.
(PDF)

S2 Fig. Tree-plot for a single cell (left) and fixed surprise model prediction (right). Each column of the tree-plot shows the average response of the neuron to all stimulus sequences of a given length, that end with flash.
(PDF)

S3 Fig. Tree-plot for three representative cells, showing neuron's response (left column), and prediction by the adaptive surprise model (right column).
(PDF)

S4 Fig. Tree-plot for reduced adaptive surprise model, showing neuron's response (left column), and prediction by the adaptive surprise model (right column). The qualitative traits of the adaptive surprise model remain even if two instead of four prior parameters are used.
(PDF)

S5 Fig. Dynamic inference model performs similar to its leaky integration approximation. A. Stimulus excerpt (above) and recorded PSTH (below, black), and prediction of the dynamic

surprise model (below, red). **B.** Pearson correlation coefficients of model fits to each cell's PSTH, for the dynamic surprise model (x-axis) versus the adaptive surprise model (y-axis). Two model perform similarly (the correlation coefficients for the respective model fits lie close to the unity line) despite the significant difference in mean ($p = 3 \cdot 10^{-5}$, Wilcoxon signed-rank test). As such, the adaptive surprise model can be treated as an approximation of the dynamic surprise model. **C.** Tree-plot, showing the mean response of two representative cells to different sequences of flashes (filled circles) and silences (empty circles). Each column of the tree-plot shows the average response of the neuron to all stimulus sequences of a given length, that end with silence (top) or flash (bottom).
(PDF)

S6 Fig. Adaptive surprise model for a repeated experiment. A. Stimulus excerpt (above) and recorded PSTH (below, black), alongside prediction of the adaptive surprise model (below, blue). **B.** Neural responses to varying number of consecutive flashes (above). Recorded PSTH of three neurons is shown to the left, while model predictions are shown to the right. Each colour corresponds to a different number of consecutive flashes. The adaptive surprise model captures variations in both the magnitude and width of the OSR. **C.** Increase in the OSR with the number of consecutive flashes for seven cells (each cell plotted with a different colour). The data (solid line) is plotted alongside the predictions of the adaptive surprise model (solid lines with circles). **D.** Pearson correlation coefficients between each cell's PSTH and the model predictions, for the fixed surprise model (y-axis) versus the adaptive surprise model (x-axis). The adaptive surprise model significantly outperforms the fixed surprise model ($p = 1 \cdot 7^{-20}$, Wilcoxon signed-rank test).
(PDF)

S7 Fig. Cell typing for repeated experiment. (Top left panel) Full-field 'chirp' stimulus used to categorise different cell-types in the repeat experiment. (Lower left panels) Mean (red) and standard deviation (gray) of PSTH in response to the chirp stimulus after clustering into 8 different cell-types. (Right panels) Average temporal STA in response to binary checkerboard stimulus. Each categorised cell-type was further classified manually into putative ON, OFF & ON-OFF types, depending on their response to a prolonged on and off flash, at the start of the chirp stimulus.
(PDF)

S8 Fig. Putative cell types in the second experiment, based on responses to chirp stimulus. For three identified cell types (ON, OFF, ON-OFF), we show the number of cells with/without an OSR.
(PDF)

S9 Fig. Distribution of the fitted prior depending on the cell type. Each dot represents a cell. Cells have been clustered as in S7 Fig. We have also indicated how each of the broad groups divides into the ON, OFF, ON-OFF categories.
(PDF)

S10 Fig. Tree history: Individual examples and population average. A. We wanted to quantify how neural responses depended on stimuli going back in the past. To do this, we quantified the 'branching distance' at each depth of the tree-plot, defined as the difference between branches that extending from the same 'parent' node. Plotted here are the results for the three cells shown in Fig 4. The adaptive surprise model was able to capture the qualitative shape of this curve for these cells. **B.** Population average of the branching distance (bars represent standard error). The main difference between data and model was that the branching distance

decreases to zero for the model as the tree depth increased, unlike the data. This is likely due to the noise in the empirical estimates of firing rate from data, which results in small random variations in the positions of the branches in the tree plot.
(PDF)

S11 Fig. A. Response of a model cell, using the model of Werner et al. [13]. **B.** Omitted stimulus response versus number of consecutive flashes, for one cell predicted by the Werner model. **C.** When we fitted the 3 free parameters of the model to our data, it was not able to fit the responses as well as the adaptive surprise model, with signicantly lower correlation coefficient ($p = 2 \cdot 10^{-9}$, Wilcoxon signed-rank test).
(PDF)

S12 Fig. Learned priors over the transition probability, $\theta = p(x_t = 1|x_{t-1})$, fitted to data. Here we plot a subset of 6 cells.
(PDF)

S1 Text. Dynamic surprise model.
(PDF)

S2 Text. Cell typing.
(PDF)

Acknowledgments

We would like to thank Ulisse Ferrari, Francesco Trapani, Matias Goldin and Samuele Virgili for useful discussions.

Author Contributions

Conceptualization: Danica Despotović, Olivier Marre, Matthew Chalk.

Funding acquisition: Matthew Chalk.

Investigation: Danica Despotović, Corentin Joffrois, Matthew Chalk.

Methodology: Danica Despotović, Corentin Joffrois, Matthew Chalk.

Project administration: Olivier Marre, Matthew Chalk.

Resources: Olivier Marre.

Software: Danica Despotović, Matthew Chalk.

Supervision: Olivier Marre, Matthew Chalk.

Visualization: Danica Despotović.

Writing – original draft: Danica Despotović, Matthew Chalk.

Writing – review & editing: Danica Despotović, Olivier Marre, Matthew Chalk.

References

1. Karklin Y & Simoncelli E. Efficient coding of natural images with a population of noisy linear-nonlinear neurons. Advances In Neural Information Processing Systems. 2011: 24 PMID: 26273180
2. Doi E, Gauthier J, Field G, Shlens J, Sher A, Greschner M, Machado T, Jepson L, Mathieson K, Gunning D & Litke AM. Efficient coding of spatial information in the primate retina. Journal Of Neuroscience. 2012; 32(46):16256–16264 (2012) https://doi.org/10.1523/JNEUROSCI.4036-12.2012 PMID: 23152609

3. Soto F, Hsiang J, Rajagopal R, Piggott K, Harocopos G, Couch S, Custer P, Morgan J & Kerschensteiner D. Efficient coding by midget and parasol ganglion cells in the human retina. Neuron. 2020; 107 (4):656–666 https://doi.org/10.1016/j.neuron.2020.05.030 PMID: 32533915
4. Rao RP, Ballard DH. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature neuroscience. 1999; 2(1):79–87. https://doi.org/10.1038/4580 PMID: 10195184
5. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381(6583):607–9. https://doi.org/10.1038/381607a0 PMID: 8637596
6. Van Hateren JH, van der Schaaf A. Independent component filters of natural images compared with simple cells in primary visual cortex. Proceedings of the Royal Society of London. Series B: Biological Sciences. 1998; 265(1394):359–66. https://doi.org/10.1098/rspb.1998.0303 PMID: 9523437
7. Lewicki MS. Efficient coding of natural sounds. Nature neuroscience. 2002; 5(4):356–63. https://doi.org/10.1038/nn831 PMID: 11896400
8. Smith EC, Lewicki MS. Efficient auditory coding. Nature. 2006; 439(7079):978–82. https://doi.org/10.1038/nature04485 PMID: 16495999
9. Ulanovsky N, Las L, Nelken I. Processing of low-probability sounds by cortical neurons. Nature neuroscience. 2003; 6(4):391–8. https://doi.org/10.1038/nn1032 PMID: 12652303
10. Gill P, Woolley SM, Fremouw T, Theunissen FE. What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. Journal of neurophysiology. 2008; 99(6):2809–20. https://doi.org/10.1152/jn.01270.2007 PMID: 18287545
11. Schwartz G, Harris R, Shrom D, Berry MJ. Detection and prediction of periodic patterns by the retina. Nature neuroscience. 2007; 10(5):552–4. https://doi.org/10.1038/nn1887 PMID: 17450138
12. Schwartz G, Berry MJ 2nd. Sophisticated temporal pattern recognition in retinal ganglion cells. Journal of neurophysiology. 2008; 99(4):1787–98. https://doi.org/10.1152/jn.01025.2007 PMID: 18272878
13. Werner B, Cook PB, Passaglia CL. Complex temporal response patterns with a simple retinal circuit. Journal of neurophysiology. 2008; 100(2):1087–97. https://doi.org/10.1152/jn.90527.2008 PMID: 18579656
14. Marre O, Amodei D, Deshmukh N, Sadeghi K, Soo F, Holy TE, Berry MJ. Mapping a complete neural population in the retina. Journal of Neuroscience. 2012; 32(43):14859–73. https://doi.org/10.1523/JNEUROSCI.0723-12.2012 PMID: 23100409
15. Yger P, Spampinato GL, Esposito E, Lefebvre B, Deny S, Gardella C, Stimberg M, Jetter F, Zeck G, Picaud S, Duebel J. Fast and accurate spike sorting in vitro and in vivo for up to thousands of electrodes. BioRxiv. 2016:067843.
16. Meyniel F, Maheu M, Dehaene S. Human inferences about sequences: A minimal transition probability model. PLoS computational biology. 2016; 12(12):e1005260. https://doi.org/10.1371/journal.pcbi.1005260 PMID: 28030543
17. Doya K, editor. Bayesian brain: Probabilistic approaches to neural coding. MIT press; 2007.
18. Shannon CE. A mathematical theory of communication. The Bell system technical journal. 1948; 27 (3):379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x
19. MacKay DJ. Information theory, inference and learning algorithms. Cambridge university press; 2003
20. Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. Journal of Neuroscience. 2005; 25(47):11003–13. https://doi.org/10.1523/JNEUROSCI.3305-05.2005 PMID: 16306413
21. Geisler WS. Sequential ideal-observer analysis of visual discriminations. Psychological review. 1989 Apr; 96(2):267. https://doi.org/10.1037/0033-295X.96.2.267 PMID: 2652171
22. Smeds L, Takeshita D, Turunen T, Tiihonen J, Westo J, Martyniuk N, Seppanen A, Ala-Laurila P. Paradoxical rules of spike train decoding revealed at the sensitivity limit of vision. Neuron. 2019; 104(3):576–87. https://doi.org/10.1016/j.neuron.2019.08.005 PMID: 31519460
23. Chichilnisky EJ, Rieke F. Detection sensitivity and temporal resolution of visual signals near absolute threshold in the salamander retina. Journal of Neuroscience. 2005; 25(2):318–30. https://doi.org/10.1523/JNEUROSCI.2339-04.2005 PMID: 15647475
24. Baden T, Berens P, Franke K, Román Rosón M, Bethge M, Euler T. The functional diversity of retinal ganglion cells in the mouse. Nature. 2016; 529(7586):345–50. https://doi.org/10.1038/nature16468 PMID: 26735013
25. Neitz J, Carroll J, Yamauchi Y, Neitz M, Williams DR. Color perception is mediated by a plastic neural mechanism that is adjustable in adults. Neuron. 2002; 35(4):783–92. https://doi.org/10.1016/S0896-6273(02)00818-8 PMID: 12194876

26. Chalk M, Seitz AR, Series P. Rapidly learned stimulus expectations alter perception of motion. Journal of vision. 2010; 10(8) https://doi.org/10.1167/10.8.2 PMID: 20884577

27. Hosoya T, Baccus SA, Meister M. Dynamic predictive coding by the retina. Nature. 2005; 436 (7047):71–7. https://doi.org/10.1038/nature03689 PMID: 16001064

28. Wark B, Fairhall A, Rieke F. Timescales of inference in visual adaptation. Neuron. 2009; 61(5):750–61. https://doi.org/10.1016/j.neuron.2009.01.019 PMID: 19285471

29. Mlynarski WF, Hermundstad AM. Adaptive coding for dynamic sensory inference. Elife. 2018; 7: e32055. https://doi.org/10.7554/eLife.32055 PMID: 29988020

30. Mlynarski WF, Hermundstad AM. Efficient and adaptive sensory codes. Nature Neuroscience. 2021; 24 (7):998–1009. https://doi.org/10.1038/s41593-021-00846-0 PMID: 34017131

31. Ebert S, Buffet T, Sermet BS, Marre O, Cessac B. Temporal pattern recognition in retinal ganglion cells is mediated by dynamical inhibitory synapses. bioRxiv. 2023:2023–01.

32. Deshmukh NR. Complex computation in the retina. Diss. Princeton University, 2015

33. Maheswaranathan N, McIntosh LT, Tanaka H, Grant S, Kastner DB, Melander JB, Nayebi A, Brezovec LE, Wang JH, Ganguli S, Baccus SA. Interpreting the retinal neural code for natural scenes: From computations to neurons. Neuron. 2023; 111(17):2742–55. https://doi.org/10.1016/j.neuron.2023.06.007 PMID: 37451264

34. Tanaka H, Nayebi A, Maheswaranathan N, McIntosh L, Baccus S, Ganguli S. From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. Advances in neural information processing systems. 2019; 32. PMID: 35283616

35. Chen KS, Chen CC, Chan CK. Characterization of predictive behavior of a retina by mutual information. Frontiers in computational neuroscience. 2017; 11:66. https://doi.org/10.3389/fncom.2017.00066 PMID: 28775686

36. Keller GB, Bonhoeffer T, Hubener M. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. Neuron. 2012; 74(5):809–15. https://doi.org/10.1016/j.neuron.2012.03.040 PMID: 22681686

37. Machens CK, Gollisch T, Kolesnikova O, Herz AV. Testing the efficiency of sensory coding with optimal stimulus ensembles. Neuron. 2005; 47(3):447–56 https://doi.org/10.1016/j.neuron.2005.06.015 PMID: 16055067