

Specificity and timescales of cortical adaptation as inferences about natural movie statistics

Michael Snow

Department of Systems and Computational Biology,
and Dominick Purpura Department of Neuroscience,
Albert Einstein College of Medicine, Bronx, NY, USA



Ruben Coen-Cagli

Department of Basic Neuroscience, University of Geneva,
Switzerland Department of Systems and Computational Biology,
and Dominick Purpura Department of Neuroscience,
Albert Einstein College of Medicine, Bronx, NY, USA



Odelia Schwartz

Department of Computer Science, University of Miami,
Miami, FL, USA
Dominick Purpura Department of Neuroscience,
and Department of Systems and Computational Biology,
Albert Einstein College of Medicine, Bronx, NY, USA



Adaptation is a phenomenological umbrella term under which a variety of temporal contextual effects are grouped. Previous models have shown that some aspects of visual adaptation reflect optimal processing of dynamic visual inputs, suggesting that adaptation should be tuned to the properties of natural visual inputs. However, the link between natural dynamic inputs and adaptation is poorly understood. Here, we extend a previously developed Bayesian modeling framework for spatial contextual effects to the temporal domain. The model learns temporal statistical regularities of natural movies and links these statistics to adaptation in primary visual cortex via divisive normalization, a ubiquitous neural computation. In particular, the model divisively normalizes the present visual input by the past visual inputs only to the degree that these are inferred to be statistically dependent. We show that this flexible form of normalization reproduces classical findings on how brief adaptation affects neuronal selectivity. Furthermore, prior knowledge acquired by the Bayesian model from natural movies can be modified by prolonged exposure to novel visual stimuli. We show that this updating can explain classical results on contrast adaptation. We also simulate the recent finding that adaptation maintains population homeostasis, namely, a balanced level of activity across a population of neurons with different orientation preferences. Consistent with previous disparate observations, our work further clarifies the

influence of stimulus-specific and neuronal-specific normalization signals in adaptation.

Introduction

Both perception and neural processing of a stimulus are dramatically influenced by what has been observed in the past, its temporal context. Such influences, broadly called *adaptation* (Carandini et al., 2005; Clifford et al., 2007; Kohn, 2007; Solomon & Kohn, 2014; Webster, 2011), have been a topic of fascination at least since the time of Aristotle, as revealed by perceptual aftereffects (Clifford & Rhodes, 2005; Schwartz, Hsu, & Dayan, 2007). Adaptation has been observed in many areas of the brain, such as the visual (Kohn, 2007), auditory (Pérez-González & Malmierca, 2014), olfactory (Kurahashi & Menini, 1997; Wilson, 2009), and somatosensory (Maravall, Petersen, Fairhall, Arabzadeh, & Diamond, 2007) regions. In the natural environment, sensory signals are always embedded in a temporal context, and correct inferences about the perceptual identity and behavioral relevance of the signals depend heavily on such context. This context-dependent inference has led to the proposal that adaptation is a hallmark of systems optimized to the temporal structure of the natural environment (e.g., Barlow & Földiák, 1989; Dayan, Sahani, & Deback,

Citation: Snow, M., Coen-Cagli, R., & Schwartz, O. (2016). Specificity and timescales of cortical adaptation as inferences about natural movie statistics. *Journal of Vision*, 16(13):1, 1–27, doi:10.1167/16.13.1.



2002; Fairhall, Lewen, Bialek, & de Ruyter Van Steveninck, 2001; Lochmann, Ernst, & Deneve, 2012; Schwartz et al., 2007; Wainwright, Schwartz, & Simoncelli, 2002; see also Attneave, 1954; Barlow, 1961).

Here we address primary visual cortex (V1) as a paradigmatic example and focus on orientation adaptation phenomena that are within the classical receptive field (RF). Over the past decades, visual cortical adaptation has been studied in great detail. Early studies revealed that the responses of neurons can be altered and often reduced over time in response to prolonged stimulation (Maffei, Fiorentini, & Bisti, 1973; Movshon & Lennie, 1979; Vautin & Berkley, 1977). Adaptation to oriented stimuli reveals a host of effects that range from contrast adaptation (Albrecht, Farrar, & Hamilton, 1984; Bonds, 1991; Ohzawa, Sclar, & Freeman, 1982, 1985; Sclar, Lennie, & DePriest, 1989), to suppression and repulsion of tuning curves in single neurons (Dragoi, Sharma, & Sur, 2000; Felsen et al., 2002; Müller, Metha, Krauskopf, & Lennie, 1999; Patterson, Wissig, & Kohn, 2013), to the recent finding that adaptation counteracts small biases in the stimulus ensemble to achieve homeostasis or equalization of population responses (Benucci, Saleem, & Carandini, 2013). Adaptation is further multifaceted in that it is triggered by some features but not others (termed *stimulus specificity*; Solomon & Kohn, 2014), and its strength often depends on neuronal selectivity (Benucci et al., 2013). Moreover, adaptation operates at a range of timescales, from milliseconds to seconds and minutes, or even longer (Dragoi et al., 2000; Kohn, 2007; Patterson et al., 2013; Solomon & Kohn, 2014). A more comprehensive treatment of experimental adaptation effects in V1 is discussed in the Introduction to V1 Adaptation Experimental Literature section.

Despite this rich experimental characterization, there is currently not a single modeling framework that encompasses this range of effects. A long-standing proposal is that divisive normalization, a unifying descriptive model for many aspects of signaling in sensory areas, could naturally account for adaptation (Albrecht & Geisler, 1991; Carandini & Heeger, 2012; Heeger, 1992). However, descriptive models do not determine when adaptation should occur or the parameters that govern its strength, which is crucial for capturing the range of experimental data we consider here. A complementary approach is to consider theoretical frameworks that posit that adaptation can be explained as optimal processing of temporal signals (Buiatti & van Vreeswijk, 2003; Lochmann et al., 2012; Stevenson, Cronin, Sur, & Kording, 2010; Stocker & Simoncelli, 2006; Wainwright et al., 2002). These studies rely, in different ways, on the hypothesis that

sensory processing is matched to the statistical properties of natural scenes (Barlow, 1961; Olshausen & Field, 2000) and that it is a form of Bayesian inference based on a generative model, that is, a probabilistic relationship between objects in the environment and the sensory inputs they produce (Barlow, 2001; Berkes, Orbán, Lengyel, & Fiser, 2011; Dayan & Abbott, 2001; Knill & Richards, 1996; Stocker & Simoncelli, 2006). However, none of these recent studies addressed a crucial requirement of this approach, namely, that such a model be learned from the temporal statistics of natural movies to form, and update, a Bayesian prior about temporal expectations in the natural environment.

To address these limitations, we took a different approach. We hypothesized that cortical neurons are sensitive to temporal regularities inherent in the natural visual environment. We constructed a model that learns such prior knowledge and performs Bayesian inference on the visual input based on that prior knowledge. Because objects in the world generally are either static (e.g., a tree, a rock) or move slowly and smoothly (e.g., a person walking on the beach; Weiss & Adelson, 1998), the inputs to V1 neurons change little over short timescales (e.g., within one eye fixation). Therefore, present and past visual inputs are often highly redundant. Across longer timescales, or in contexts with quickly changing motion, such as a carousel, temporal sequences of visual inputs are less redundant (Figure 1, lion vs. carousel). Our model developed similar prior knowledge through exposure to an ensemble of natural movies and learned to judge the statistical similarity between the past and the present in new inputs. We thus developed a Bayesian modeling framework that learned regularities in natural movies over time and then sought to reproduce some adaptation effects in V1.

The framework is closely related to divisive normalization, whereby in the temporal version, the response of a neuron is divisively normalized by the responses of other neurons across time (Heeger, 1992). However, the Bayesian framework extends the standard normalization model in a critical way; responses are normalized only to the degree that past and present inputs are inferred to be statistically similar. We first optimized a set of parameters in the generative model that characterize the dependencies between movie frames over hundreds of milliseconds. This approach captured some neurophysiology data for short-term adaptation. Furthermore, the Bayesian model could be naturally extended to capture longer timescales, on the order of seconds, by updating the parameters that govern the generative model. This allowed the model to also reflect longer-term statistics of the experimental stimulus ensemble and to capture a broader range of experimental data.

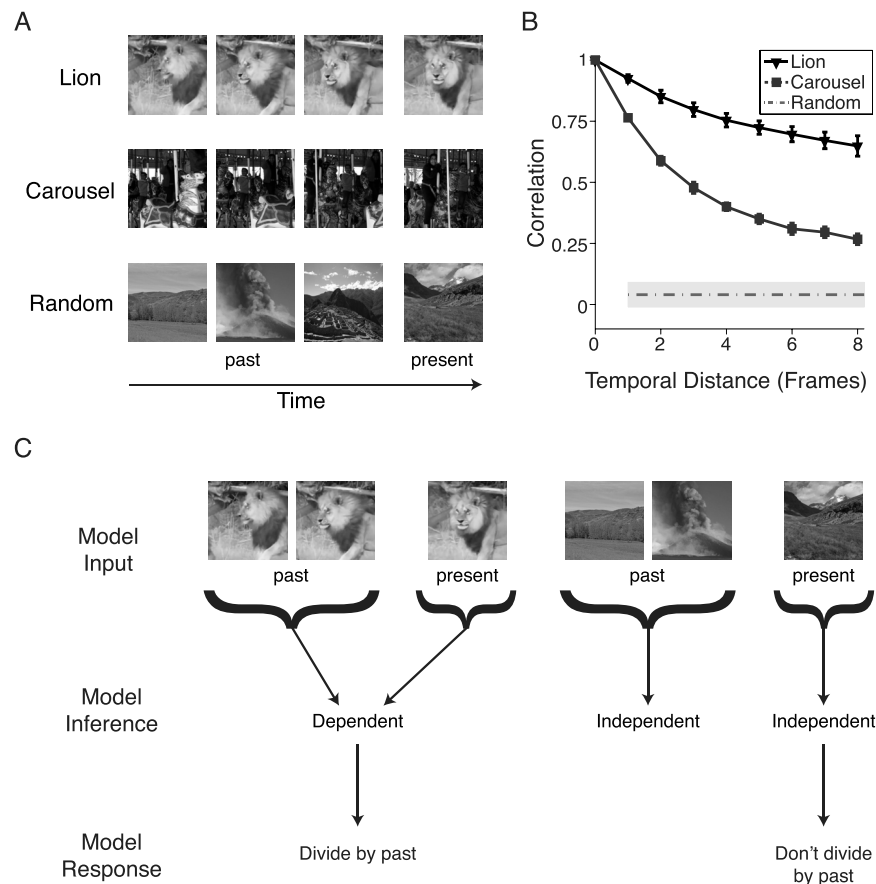


Figure 1. Dependencies in natural movies and a generalized divisive normalization model for reducing dependencies. (A) Past (left columns) and present (right column) frames of natural movies (top two rows) versus a movie composed of random still natural images (bottom row). (B) Filled symbols: Correlation coefficient of the responses of a standard complex cell model to the movies in (A), namely, the square root of sum of squared responses of two quadrature pair-oriented filters. The abscissa is the temporal distance used for the calculation of correlation between the “present” and “past” frames. The error bars are ± 1 SD across different filter orientations. Dashed line and shaded area: Average correlation coefficient ± 1 SD over all correlations for all distances, because the order of the random frames is artificial, for the movie composed of random still images. (C) Flexible divisive normalization. The model divides the present by the past if the past and present share dependencies (left case) and does not divide the present by the past when the past and present are independent (right case). Lion images adapted from “Be Bold as a Lion” by Kids on the Move, <https://vimeo.com/146515611>, available under the Creative Commons Attribution license.

In designing our model, we found it useful to distinguish between two types of normalization, as has been done previously in the adaptation literature (e.g., Benucci et al., 2013; Solomon & Kohn, 2014): (a) stimulus-specific (i.e., strongest normalization when the past adapting stimulus is similar to the present test stimulus) and (b) neuronal-specific (i.e., strongest normalization when the adapting stimulus matches the preference of the neuron). Each model implementation captured a subset of adaptation effects, suggesting that these represent slices through a more complicated space. Indeed, depending on the adaptation conditions, both effects are likely present to different degrees.

Overall, this work provides a step forward toward a common framework for understanding a diverse set of visual adaptation phenomena.

Introduction to V1 adaptation experimental literature

We provide a brief summary of adaptation neurophysiology in V1 and then specify the effects that we seek to address with our modeling approach. We focus specifically on orientation adaptation effects in V1. Other factors such as spatial and temporal frequency are also known to influence adaptation (e.g., King, Lowe, & Crowder, 2015; Ledue, King, Stover, & Crowder, 2013; Saul & Cynader, 1989) but are beyond the scope of our current modeling effort, which is based on a population of oriented RFs. We also focus on effects confined to the classical RF of V1 neurons. A more comprehensive review of adaptation effects in V1 and other areas can be found in several review papers

(Clifford & Rhodes, 2005; Clifford et al., 2007; Ibbotson, 2005; Kohn, 2007; Krekelberg, Boynton, & van Wezel, 2006; Schwartz et al., 2007; Solomon & Kohn, 2014).

A main property of V1 adaptation that has been documented extensively is the reduced response to test stimuli following adaptation. When the test stimulus is matched in orientation to the RF of a given neuron, most suppression is typically found when the adapting stimulus also matches the neuron's preferred RF (Carandini, Movshon, & Ferster, 1998; Crowder et al., 2006; Nelson, 1991). Adaptation to a given stimulus is often depicted in terms of changes in tuning curves. For an adapter optimally matched to the classical RF, multiple labs have found response suppression of tuning curves (e.g., Crowder et al., 2006; Dragoi et al., 2002; Felsen et al., 2002; Müller et al., 1999).

An interesting facet of adaptation is that when one adapts on the flank of the tuning curve, this often results in a repulsive shift of the tuning curve. This effect is striking and has been observed across multiple labs (Dragoi et al., 2000; Felsen et al., 2002; Müller et al., 1999; Patterson et al., 2013), even when adapting for short timescales of hundreds of milliseconds (Dragoi, Sharma, Miller, & Sur, 2002; Müller et al., 1999). This is the classical example of the so-called stimulus-specific adaptation, because the amount of suppression is maximal when the adapter is matched to the test, not when it is matched to the neuron's preferred RF. This distinction cannot be made for the aforementioned case of tuning curve suppression, because the maximal suppressive effect is observed only when the adapter is matched to both the test stimulus and the preferred orientation of the neuron. The observation of stimulus specificity also applies to other stimulus attributes, as Movshon and Lennie (1979, p. 852) noted in earlier work with spatial frequency: "Our most surprising observation is that the loss of sensitivity in cortical neurones can be specific to the adapting stimulus."

Adaptation is influenced by timescale. Effects of tuning curve changes, such as suppression and repulsion, are typically increased for longer adaptation times (Dragoi et al., 2000; Patterson et al., 2013), although there are exceptions to such duration scaling (Patterson et al., 2013; Solomon & Kohn, 2014).

Adaptation, however, can sometimes enhance responsiveness and lead to attractive shifts in tuning curves. This has been observed in the context of surround influences and disinhibition (Solomon & Kohn, 2014; Webb, Dhruv, Solomon, Tailby, & Lennie, 2005; Wissig & Kohn, 2012). Enhancement and attraction have also been documented in area MT; however, it is interesting to note that disparities between V1 and MT on such fronts have also been attributed to stimulus size (Patterson, Duijnhouwer, Wissig, Krekelberg, &

Kohn, 2014; Wissig & Kohn, 2012). Some studies have suggested attraction in V1 for long adaptation periods of minutes (Jeyabalaratnam et al., 2013), but these also included stimuli that were much larger than the classical RF. We do not have surround influences in our adaptation model and so focus only on effects that pertain to the classical RF.

Unlike adapters that are equal to the preferred orientation or on the flank of the tuning curve, orthogonal adapters have led to more variable results. They may, in some cases, result in an increased response for test stimuli (see Dragoi et al., 2002, for short timescales). Enhancement is possibly due to recovery from reduced responsiveness to previous adaptation (see discussion in Solomon & Kohn, 2014). But in some cases, orthogonal adapters have resulted in suppression (Crowder et al., 2006), although this effect is usually weaker than adaptation to the preferred orientation (Carandini et al., 1998).

Another factor that has been studied extensively in adaptation is the influence of contrast. Contrast adaptation causes a change in contrast sensitivity via a rightward shift of the contrast response curve (with larger shifts for adaptation to higher contrasts) and may also change its saturation level (Albrecht et al., 1984; Bonds, 1991; Ohzawa et al., 1982, 1985; Sclar et al., 1989).

In this article, we first show that we can replicate some adaptation effects, which have been described in multiple labs, namely, suppression and repulsion of tuning curves. We explain how these effects are natural consequences of our model, even on short timescales, because of its learning of regularities in movie statistics. Normalization occurs in our framework only to the degree that past and present are deemed statistically dependent. We then consider other experimental phenomena. We examine contrast adaptation and shifts in tuning curves, which have also been documented in many labs. We show that, in line with other normalization frameworks, we can capture such contrast shifts. This is captured by our model, provided that we update our model during the adaptation, thereby reflecting an adapting period on the order of seconds. We then focus on an interesting recent phenomenon, response equalization, documented by Benucci et al. (2013). In this paradigm, multiple orientations are presented for a prolonged period of adaptation over several seconds. Although one orientation (the biased orientation) is presented more frequently than other orientations, the responses of neurons across the population are equalized following adaptation, reflecting a form of homeostasis. We show that these data could not be explained by the same model that captured stimulus-specific suppression. Modeling response equalization required a variant of the model, which includes neuronal-specific suppression. Finally,

in the Discussion section, we consider how to apply our approach to other longer-term adaptation studies that have found strengthening of tuning curve suppression and repulsion for longer adaptation durations. We also discuss limitations of our approach.

Methods

Overview of modeling framework

We consider a framework closely related to divisive normalization models of adaptation. In the standard view of normalization, the response of a neuron to a sequence of images is obtained by applying an oriented filter and then dividing by a combination of the outputs of other filters at different times and orientations. Here we link this operation to the statistics of natural movies and consider several extensions.

Following a large body of work on image statistics, we hypothesize that the firing rates of different V1 neurons represent the intensity of features that may be present in the visual input, such as oriented Gabors. In other words, given the visual input, a neuron computes the intensity of the feature it represents. This cannot be achieved simply by linear filtering; for instance, the output of a vertical filter applied to a vertical, low-contrast image may be the same as when it is applied to a diagonal, high-contrast image. In this example, division by the sum of squares of the outputs of filters with different orientations (i.e., a measure of contrast) disambiguates the two cases.

In general, natural visual inputs are much more complex, as they combine multiple features of different contrasts across time and space. Here we focus specifically on the temporal domain. Therefore, to compute feature intensity, neurons need to rely on some knowledge about the structure of images across time, to correctly infer the intensity. In studies of image statistics, this knowledge is typically formalized by a generative model of images, namely, a model that specifies how likely individual features and contrasts are, a priori, and how to combine them. Inference then requires inverting the generative model (i.e., applying Bayes' rule to compute a posterior probability over feature intensity given the visual input).

This article is based on a class of generative models known as Gaussian scale mixtures (GSM; Andrews & Mallows, 1974; Wainwright & Simoncelli, 2000) and extensions to mixtures of GSM (MGSM; Coen-Cagli, Dayan, & Schwartz, 2009; Schwartz, Sejnowski, & Dayan, 2006). Our previous work showed that these models are closely related to divisive normalization and account for some spatial context phenomena (Coen-Cagli, Dayan, & Schwartz, 2012; Coen-Cagli et al.,

2009; Coen-Cagli & Schwartz, 2013; Schwartz, Sejnowski, & Dayan, 2009). The models also relate to ideas of breakdown of spatial homogeneity and salience in V1 (Li, 1999). Here we apply the MGSM modeling framework to temporal stimuli (i.e., movie frames) and neural adaptation effects.

Before describing the models in detail, we provide here the intuition behind the temporal GSM and MGSM. Subsequent frames of a natural movie are highly redundant, as objects tend to move smoothly and slowly, and therefore the outputs of filters applied to subsequent frames are statistically dependent. Such dependencies are modeled in the GSM as arising from the multiplication between a set of local variables (analogous to feature intensities across time and orientations) and a global variable (analogous to contrast). Given an input movie and the corresponding filters' outputs, the model neural response (i.e., an estimate of the local intensity variable) is obtained by estimating the contrast variable and dividing it out, a form of divisive normalization. This model, however, cannot account for the fact that the dependencies may be stronger between some features than others or absent altogether (e.g., when there is fast motion through the visual field and subsequent frames contain different objects). This is captured by MGSMs, in which inference proceeds in two steps: Given an input movie, one first infers which subset of features, if any, are statistically dependent; second, one then divides the target filter's output by a normalization signal computed only from the other filters that were inferred dependent with it.

Below, we briefly describe the linear filters we used to quantify how strongly past and present frames of a movie drive both the RF and the normalization pool of model neurons. This is followed by an overview of the general formulation for this type of model and the link to divisive normalization. We then describe two different implementations of the model for adaptation in V1. The two implementations differ by how they capture the statistical similarity (i.e., the dependence) between the past and present, namely, whether it is stimulus specific or neuronal specific. We then briefly enumerate the process by which we trained the models. Finally, we conclude by extending the models to address temporal dynamics over longer timescales.

V1 model RFs

We used a bank of oriented visual filters, qualitatively similar to V1 RFs, to compute both the excitatory drive to the RF of a neuron and the normalization signal. Specifically, the filters were drawn from the first level of a steerable pyramid (Simoncelli, Freeman, Adelson, & Heeger, 1992), with a diameter of 9 pixels and a peak spatial frequency of

1/6 cycles/pixels. We considered filters with four possible orientations (0° , 45° , 90° , and 135°), denoted by θ , along with two possible phases (0° and 90° to form a quadrature pair) and nine temporal positions (eight past and one present temporal position). Using movies with frame rates of 30 frames per second (fps), we chose an adapt period of eight frames, which corresponds to about 240 ms, because this is within the range of standard short-term adaptation experimental paradigms (Kohn, 2007). Below, we further discuss model extensions for adaptation over longer timescales.

For a given image, the output of a filter, from here on simply referred to as the RF output, is given by the inner product between the image and the filter. We generically denote the vector of present RF outputs by x_t and the past RF outputs by x_p . We considered different choices for x_t and x_p , as detailed below for the different versions of the model (see the sections titled “Binary MGSM and Flexible Divisive Normalization” and “Multiorientation MGSM Model”).

Standard Gaussian scale mixture model and divisive normalization

A general property of natural movies is that for a single spatiotemporal position, neighboring pixels in the spatial and temporal domains are highly redundant (Figure 1A). The spatial redundancy is due to global properties of the image, such as contrast, which are shared between neighboring areas. Because movies change slowly and smoothly over time (Weiss & Adelson, 1998), neighboring temporal positions (e.g., in the past and present time frames) are highly redundant as well (e.g., Dong & Atick, 1995). The GSM (Andrews & Mallows, 1974; Wainwright & Simoncelli, 2000) is a generative model that captures these dependencies (i.e., generates RF outputs that have the typical dependencies observed in natural stimuli).

The GSM generates the RF outputs for the past and present frames, x_t and x_p , respectively, such that they are statistically dependent, by the multiplication of two random variables: (a) a Gaussian variable that represents structure local to each RF, which we label g_t and g_p , for the present and past RFs, respectively, and (b) a positive scalar random variable, v , which is shared between multiple RFs (and essentially captures the dependence among them). The nonlinear dependencies between RF outputs in the past and present are introduced via the multiplication of the local Gaussian variable, g , with the shared mixer variable, v . Therefore, we model the dependent RF outputs of the present time frame and the past time frames as

$$\begin{aligned} x_t &= v g_t \\ x_p &= v g_p \end{aligned} \quad (1)$$

where g_t and g_p are vectors representing the local Gaussian variable in the present and past, respectively, and are the same size as the filter outputs to which they correspond in Equation 1, that is, x_t and x_p . To connect the GSM to temporal contextual effects, we assume that the Gaussian components associated with the present frame, g_t , relate to the firing rates of V1 neurons. Estimating this Gaussian component amounts to a form of divisive normalization, by essentially inverting the multiplicative model above (thus amounting to division).

More specifically, we need to expand the notation of Equation 1 to include the preferred orientation of a given neuron. We associate each entry of the vector (which we label θ for the preferred orientation of the corresponding RF), $g_{t,\theta}$, with the firing of a neuron whose preferred orientation reflects the corresponding RF output, $x_{t,\theta}$, in V1. We focus on the Gaussian component because it represents structure local to an RF at a single point in time. In contrast, the mixer variable, v , represents the link between RF responses across time. As the GSM is a generative model, by inverting the model and using Bayes’s rule, we can estimate the value of the local Gaussian components, g_t . Given a set of stimuli, we collect the set of RF outputs in the past and present times, denoted by the vector (x_t, x_p) , and then calculate the expected value of the local Gaussian component for the present. Because the generative model is multiplicative (as in Equation 1), the inverse operation of computing the local Gaussian amounts to a form of divisive normalization. It has been shown (Schwartz & Simoncelli, 2001) that such an operation also relates to efficient coding as it reduces the statistical dependencies that, in the GSM, are due to this common mixer variable, v . The resulting divisive normalization equation is given by (see Coen-Cagli et al., 2009; Schwartz et al., 2006, for derivations):

$$E[g_{t,\theta}|x_t, x_p] \approx \frac{|x_{t,\theta}|}{\sqrt{\lambda_{tp}}} \quad (2)$$

where the numerator of Equation 2 is the rectified RF response corresponding to the preferred orientation of the given model neuron and the denominator of Equation 2 is given by λ_{tp} , which represents the normalization signal and is defined as follows:

$$\lambda_{tp} = \sqrt{(x_t, x_p)^T (\Sigma_{tp})^{-1} (x_t, x_p)} \quad (3)$$

The covariance matrix Σ_{tp} is a measure of the linear dependencies between elements of the past and present Gaussians, g_p and g_t , and intuitively may be seen as representing how strongly the features (of orientation in the past and the present) are statistically dependent in natural scenes. Equation 3 thus amounts to a weighted sum of quadratic and bilinear combinations

of the RF responses that contribute to the divisive normalization signal. If the covariance matrix is the identity matrix, then this reduces simply to a sum of squares, similar to the original formulation of divisive normalization (Heeger, 1992).

We suggest that this GSM framework can explain both neuron- and stimulus-specific adaptation effects as determined by the statistical similarity of the past and present inputs. The next two sections describe the two different metrics of similarity.

Binary MGSM and flexible divisive normalization

We introduce here one of the models we used to capture V1 adaptation effects, termed the *binary mixture of Gaussian scale mixtures* (MGSM), which is an extension to the temporal domain of techniques previously applied to spatial context (Coen-Cagli et al., 2009). The binary model will capture neuron-specific adaptation, as described below.

In the standard GSM, the shared mixer variable between RFs captures the dependencies that exist between frames of natural movies. However, when we empirically observe movie frames over time, we find that for particular movie sequences, the present frame is not always statistically dependent with the past frames, for instance, if the past and present correspond to different textures or objects (Figure 1). We model this independent case by assuming that the past and present RFs each have their own mixer variable (v_p and v_t), rather than a shared mixer:

$$\begin{aligned} x_t &= v_t g_t \\ x_p &= v_p g_p \end{aligned} \quad (4)$$

However, a richer characterization of the dependencies in natural movies should capture the possibility that some movie frames are dependent (shared mixer) whereas others are independent (no shared mixer). Technically, this is achieved by a mixture of the dependent GSM and independent GSM.

The model is termed *binary* in that for any given input, it is a linear mixture of two GSM models: one in which the past and present are deemed independent (as in Equation 4) and therefore the present is not divisively normalized by the past, and one in which the past and present are deemed dependent (as in Equation 1 and in the standard GSM model) and therefore the present is normalized by the past.

In the binary model, the estimate of the local Gaussian variable for a neuron with a preferred orientation, θ , amounts to a weighted sum of two conditions, one in which the past and present are independent (denoted ξ_1) and one in which they are dependent (denoted ξ_2). In the independent condition, the RF response in the present,

$x_{t,\theta}$, is not normalized by the past. In the dependent condition, the present is normalized by a set of past responses with an RF orientation matching the preferred orientation of the neuron θ , $x_{p,\theta}$. In addition, similar to Coen-Cagli et al. (2009), we include in the normalization pool of both the independent and dependent conditions the multiple orientations in the present. This is motivated by the strong dependence typically observed between overlapping RFs with different orientations (Schwartz & Simoncelli, 2001), and it also guarantees local contrast normalization of the model neural responses (Heeger, 1992).

The weight for the independent condition is the posterior probability that past and present are independent, and the weight for the dependent condition is the posterior probability that the past and present RF outputs are dependent. The greater the dependency between the past and the present, the greater the proportion that the second, dependent condition contributes to the calculation of the model output. Conversely, the lesser the degree of dependency, the greater the contribution of the first, independent condition. This model implements neuronal-specific suppression, as the tuning of the normalization signal is fixed to the neuron's preferred orientation (Benucci et al., 2013; Clifford, Wenderoth, & Spehar, 2000; Schwartz et al., 2009; Seriès, Stocker, & Simoncelli, 2009). We will show that the binary model can replicate neuronal population equalization (Benucci et al., 2013; Results section "Population-Level Response Equalization With Longer-Term Adaptation"). Further mathematical details for the binary MGSM are provided in Appendix A.

Multiorientation MGSM model

We introduce here the second model we used to capture V1 adaptation effects, termed the *multiorientation MGSM*, analogous to previous work in the spatial domain (Coen-Cagli et al., 2012). As its name implies, there is no longer just a binary choice of normalization by the past, on or off, but rather there are multiple past orientation pool conditions.

The excitatory drive to both the present RF and the present normalization pool are identical to the binary MGSM. However, we consider several past normalization pools and allow model neurons to choose which (if any) pool to use for a given movie. Specifically, we considered four past pools, one per orientation, each including all past RFs with the same orientation, denoted by $x_{p,\phi} = (x_{p1,\phi}, x_{p2,\phi}, \dots, x_{p8,\phi})$, with $\phi = 0^\circ, 45^\circ, 90^\circ$, and 135° . In the first condition, denoted ξ_1 , the present and past are independent similar to the binary model, and the present, $x_{t,\theta}$, is not normalized by the past. However, unlike the binary model, which has only one dependent condition, there are now multiple conditions

in which past and present may be deemed dependent, each pertaining to a different past RF orientation. These are denoted $\xi_{2,\phi}$, with ϕ again corresponding to each of the possible past orientations. Thus, there are now four dependent pools. In each of these conditions, the RF response, $x_{t,\theta}$, is normalized by the corresponding orientation in the past, $x_{p,\theta}$. In addition, in all conditions, the present is normalized by the other orientations in the present, as in the binary model.

The normalized responses are computed for each of the above conditions separately and then summed after weighting each condition by its posterior probability. For example, a movie with strong vertical structure that persists in time will produce a high probability that the present RF and the past vertical RFs are dependent. Therefore, the term whereby the present is normalized by the vertical past will have a larger weight than the terms with horizontal past or no past normalization.

We further assumed, for computational tractability, that all neurons share the same set of normalization pools and their corresponding weights. For example, the expected value for a model neuron with a vertical RF is a proportional mix of the vertical RF response normalized only by the present and not the past, normalized both by the present and the vertical past, normalized both by the present and the horizontal past, and so on, for each of the orientations ϕ . Similarly, the expected value for a model neuron with a horizontal RF is a proportional mix of the horizontal RF response normalized only by the present and not the past, normalized both by the present and the vertical past, normalized both by the present and the horizontal past, and so on. Therefore, the normalization pool for a vertical RF and a horizontal RF is shared. By sharing normalization pools, the effective tuning of the normalization signal is determined exclusively by the orientation of the stimulus shown in the present and its match to the past stimuli and not by the tuning of the neuron's RF. Hence, we term this a *stimulus-specific normalization model* (Benucci et al., 2013; Solomon & Kohn, 2014) and apply it to replicate tuning curve suppression and repulsion (see the section titled "Tuning Curve Adaptation Reflects Sensitivity to Inputs' Statistical Similarity"). Further mathematical details for the multiorientation MGSM are provided in Appendix B.

Model training

The parameters of the model are the prior probabilities that each past orientation normalizes the present (denoted ξ_{prior} ; see Appendices A and B) and the covariances Σ (see Equation 3) that parameterize the likelihood function. Parameters were optimized on an ensemble of natural movies culled from YouTube, because of the lack of a database of standardized natural movies. The stimulus ensemble consisted of

20,000 temporal sequences, each nine frames long, extracted from 100 frames of 14 natural movies with varying temporal and spatial properties, each normalized to the same range of luminance values (copies of the clips are available from the corresponding author). To find the optimal parameters, we maximized the likelihood through a generalized expectation-maximization algorithm (the approach and equations are described in Coen-Cagli et al., 2012). Briefly, in the expectation step (E-step), we estimate the posterior probabilities for ξ given the current parameter values. In the maximization step (M-step), we search for the parameter values that maximize the so-called complete data log-likelihood, namely, the expectation of $\log(p(x_t, x_p, \xi))$ under the estimated posterior over ξ . We divided the M-step into multiple steps, one for the dependent and one for each of the independent covariance matrices, and iterated repeatedly between the complete E-step and each partial M-step.

The training was performed unsupervised. As the EM algorithm is not guaranteed to find a global maximum, the training was run multiple times with randomized initial conditions. We found that the parameter values at convergence were similar across multiple runs. All initial parameters for both model types were learned on the multiorientation model. We introduced some minimal constraints on the learned parameters to ensure that the scale of the covariance matrices was similar, both across orientations and between dependent and independent conditions for each orientation.

After the training was complete, we introduced a free parameter, ω , to quantitatively match the overall suppressive strength of recorded neuronal responses, which varies widely across experiments. The free parameter, ω , scales the normalized response in all the dependent conditions relative to the independent condition.

The value of ω is the ratio of the learned model's response normalized by the relative suppression in the data set being replicated. Note that this additional parameter does not affect in any way the qualitative behavior of the models; it only sets the overall suppression strength.

Model extensions for long-term adaptation

In both versions of the MGSM, the model first learns the parameters from an ensemble of natural movies and then holds these parameters fixed and uses them to make response predictions as new experimental stimuli are presented. Whereas the response predictions change with each set of new stimuli, the prior parameters used to calculate the predictions remain static even as the stimulus history diverges farther from the training

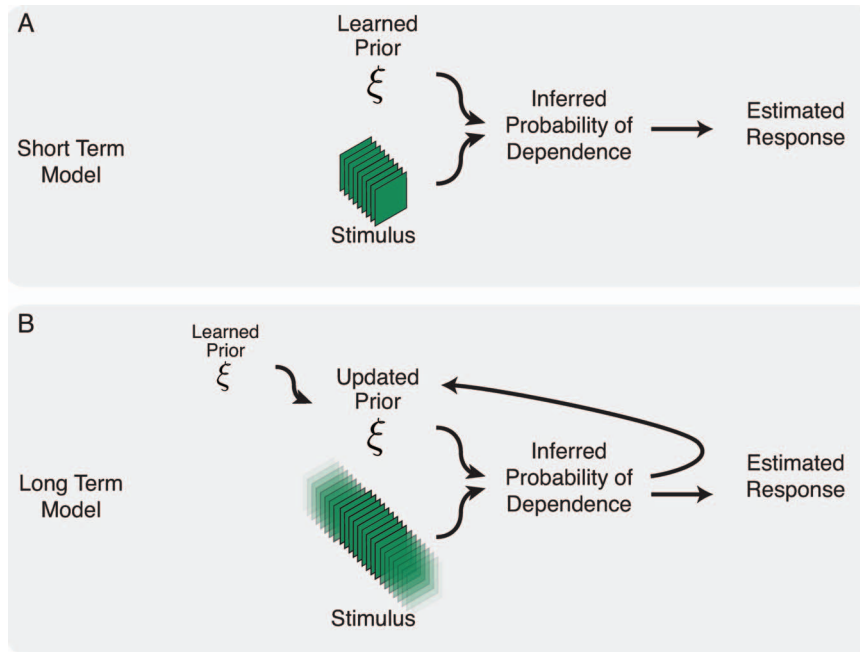


Figure 2. Schematic of model learning and inference at different timescales. (A) Short-term (static) model: The model determines the (posterior) probability of dependence between the present and past for a series of experimental stimuli, using the prior parameters learned from the movie ensemble. The prior parameters are therefore held fixed and do not change with the new experimental stimuli. The posterior probabilities in turn are used to calculate the flexible divisive normalization signal and the resulting estimated neural response. (B) Long-term (dynamic) model: The prior parameters are no longer fixed but rather updated as new experimental stimuli are presented. Only after the prior is updated does the model calculate the (posterior) dependencies inherent in a new test stimulus. The posterior probabilities are then used as in (A) to estimate the neural response.

stimulus ensemble. Given our nine time steps (eight in the past and one in the present) and a frame rate of 30 fps, the current short-term models are able to capture only effects that exist on the order of hundreds of milliseconds. Longer-term effects, which reveal themselves only on the order of seconds to minutes, would not be captured by this model.

We therefore considered a version of the models in which the prior probabilities were dynamically updated with each new set of stimulus presentations (that included eight past and one present). Specifically, for the n th movie sequence, which contains movie frames $[n-8, n-7, \dots, n]$ and RF outputs x_t^n, x_p^n , we computed the posterior $p(\xi|x_t^n, x_p^n)$ using as a prior $p(\xi|x_t^{n-1}, x_p^{n-1})$, which itself was the posterior computed for the $(n-1)$ th movie sequence. The procedure for updating was the same for each of the MGSM models (binary and multiorientation), so we describe this generically for one model. Practically, for a given movie consisting of many frames, the model was initially presented with the first set of nine frames, $[S_1, S_2, \dots, S_9]$, where S_1 through S_8 represent the past and S_9 represents the present. Using the learned prior, the model determined the posterior probability for that set of nine stimuli. The model was then presented with Frames 2 through 10 as a new set of stimuli, $[S_2, S_3, \dots, S_{10}]$, where S_2 through S_9 represent

the past and S_{10} represents the present. The calculated posterior from the previous step for Frames 1 to 9 now became the new prior, and the model once again calculated the posterior probability for Frames 2 through 10. This process was repeated as each new set of stimuli were presented, up to the last nine frames. Thus, the prior was updated recursively as each new frame of the movie was presented. A recursive Bayesian estimator has similarly been used to model multiple timescales in the retina (Wark, Fairhall, & Rieke, 2009). This long-term model is now able to track the changing visual environment on the order of seconds (Figure 2).

Results

Cortical modeling of temporal dependencies in natural movies

To test our hypothesis that adaptation can be predicted by Bayesian inference about dynamic visual stimuli, we first characterized the statistical structure of natural movies through V1-like RFs and then constructed a Bayesian model optimized to such structure.

To quantify the redundancy inherent in natural movies, we examined sequences of small patches of movie frames (Figure 1A), corresponding to the feedforward inputs into a V1 RF, and considered a standard description of a V1 complex cell, given by the square root of the sum of squares of quadrature pairs of filters (Carandini et al., 2005; Heeger, 1992). We found that responses to temporally adjacent frames were correlated. The strength of the correlation, for a given movie, decreased as a function of temporal distance between frames (Figure 1B). Furthermore, correlations decreased faster over time for sequences that included several objects sweeping fast through the RF, compared with movies containing little overall motion (Figure 1B, carousel vs. lion). These correlations are typical of dynamic signals generated at any given position in the visual field over short time intervals, such as the inputs to a V1 RF during a single fixation. Correlations were abolished for a control movie constructed from random still natural images rather than from movie frames (Figure 1B, random). The different levels of redundancy for the different movies are a signature of the nonstationary character of natural visual inputs and a general property of natural signals (Parra, Spence, & Sajda, 2001). Note that the correlations of Figure 1B represent both linear and nonlinear dependencies between RF outputs because they are computed on the squared RF outputs (see Coen-Cagli et al., 2009).

We hypothesized that neurons in V1 are sensitive to temporal dependencies and more specifically that the system reduces such dependencies by divisively normalizing neuronal responses to the present stimulus when it is redundant with past stimuli. To specify the computation required to remove the dependencies, one needs to understand how they are generated in the first place. Here, similar to recent studies of spatial contextual modulation (e.g., Coen-Cagli et al., 2012; Schwartz et al., 2009), we assumed that the dependencies in RF responses to movie sequences arise from structure that is shared between RFs across frames (e.g., due to slow fluctuations in contrast over time).

Specifically, we first considered a bank of visual RFs (Portilla & Simoncelli, 2000; Simoncelli et al., 1992, see the Methods section) covering multiple orientations at subsequent time steps. We modeled the dependencies between RFs across frames as the multiplication between a shared global variable (linking together RF responses across time) and a set of local hidden variables (each representing structure local to an RF at a single point in time; Andrews & Mallows, 1974; Wainwright & Simoncelli, 2000). As is common in such modeling approaches, we then assumed that visual cortical neurons invert the generative model (Dayan & Abbott, 2001) to estimate the local variables and so remove the redundancy induced by the global variable.

Because of the multiplicative interaction we assumed above, the inversion amounts to a form of divisive normalization of the current RF response by other RF responses (Coen-Cagli et al., 2009, 2012; see the Methods section), in this case, the recent past.

However, the assumption that there are always dependencies between RFs over time is too restrictive, as suggested by the reduced dependencies in the carousel movie (Figure 1B). In general, frames from natural movies are rarely either wholly dependent or independent but more often lie somewhere between the two extremes. A model attempting to reduce the dependencies between frames must first be able to quantify the degree of dependency between input frames and then reduce them accordingly, suggesting a flexible divisive normalization (see the Methods section). Briefly, the model response was determined in two steps: First, we computed the probability that the RF responses in the present are statistically coordinated with any group of RFs in the past; second, we divisively normalized RF responses in the present by the past RF responses only to the degree that they were inferred to be dependent (Figure 1C). Importantly, the parameters that determined the prior probability of dependence for the model were optimized to match RF dependencies in an ensemble of natural movies (see the Methods section).

To interpret what the model learns about the movie ensemble and to gain some intuition about the model behavior, it is useful to analyze the learned covariance matrices (see Equation 3). The covariance matrices reflect the typical patterns of temporal dependence in the model. Each covariance characterizes the statistics of the subset of movie sequences that are best explained by the corresponding model component. For instance, the bars in Figure 3A are the learned variances for the model component in which the present filters and *past vertical* filters are dependent. The learned variance of the present vertical filter is similar to the past vertical filters, whereas the other present orientations have lower variance. This learning reflects a form of similarity metric between past and present orientations, whereby the variance is high when the past vertical orientation matched the present vertical orientation. Similarly, a complementary pattern was learned for the model component in which present filters and *past horizontal* filters are dependent (Figure 3B). This time, the variance is high when the past horizontal orientation matched the present horizontal orientation. A similar trend occurs with respect to the covariance (off diagonal) elements of the matrix. Overall, for given experimental adaptation and test stimuli, this results in the highest probability of dependency when the past and present stimulus orientations are matched. This in turn enables the stimulus-specific orientation selectivity in the divisive normalization.

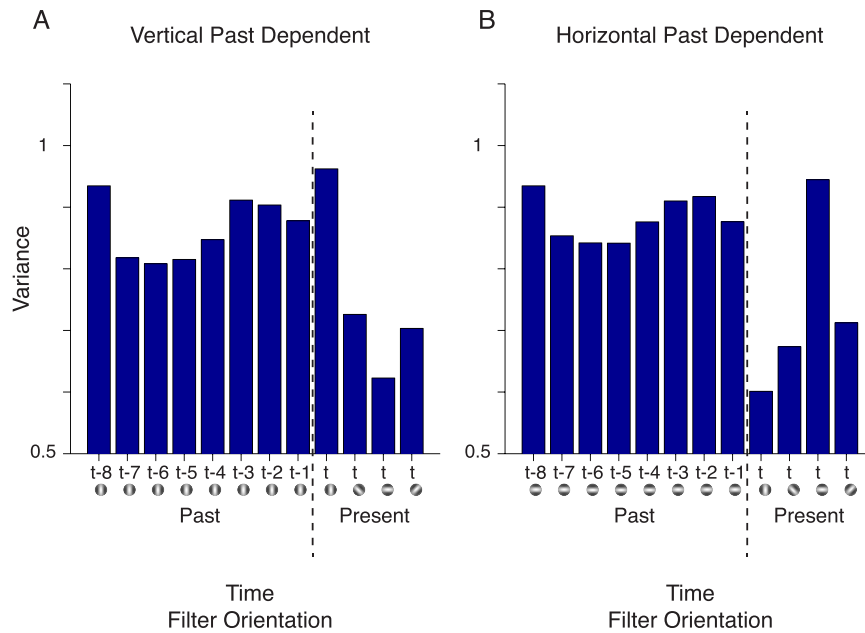


Figure 3. Learned variances in the multiorientation model. (A) Model component in which vertical past filters and present filter are dependent. The ordinate represents the learned variance of the Gaussian latent variables corresponding to a vertical filter in the past (left of the dotted line) and four filter orientations in the present (right of the dotted line). The top row of the labels on the abscissa refers to the temporal position of the frames relative to the present frame (t); for example, one frame in the past is $t-1$. The icons below each temporal position depict the filter orientation. (B) Same as for (A) but for the model component in which horizontal past filters and present filters are dependent.

Comparing model simulations to experimental data

In the next part of the Results, we simulate adaptation experiments on our model. A specific prediction of our framework is that visual adaptation results in strong divisive suppression of cortical activity only when the visual inputs are inferred statistically dependent but not otherwise. This inference relies on the prior expectation (set by exposure to the natural environment) that visual inputs are often, but not always, dependent over short time spans. Furthermore, we suggest that long-term adaptation effects could be described as a consequence of updating such prior expectations.

Tuning curve adaptation reflects sensitivity to inputs' statistical similarity

We first tested whether the specific model prediction for short-term adaptation, namely, that the effects are contingent on the inferred statistical dependency of the inputs, can account for experimental orientation adaptation data (Figure 4A, D). We specifically considered classical experiments on suppressive and repulsive changes in tuning curves that have been documented in multiple labs, even for short timescales of adaptation of hundreds of

milliseconds (Dragoi et al., 2002; Felsen et al., 2002; Müller et al., 1999; Patterson et al., 2013). Here we show that our model can replicate the main qualitative result of tuning curve suppression and repulsion. In the Discussion section, we further address studies that have found strengthening of effects with adaptation time.

We computed model responses to experimental stimuli consisting of short sequences of oriented gratings. To simulate orientation adaptation and its influence on tuning curves in the model, we considered stimulus sequences analogous to experiments on changes in tuning curves. The gratings' orientation was fixed for the past frames (adapter) and varied on each trial of the present frame (test) to measure the effects of adaptation on the neuron's tuning curve.

Figure 4A illustrates the typical effect of response suppression when an adapter is matched to the orientation preference of the neuron and the tuning curve amplitude is reduced by adaptation (experimental data plotted from Wissig & Kohn, 2012; suppression also occurs for shorter timescales of adaptation, as documented in the summary figure 2 of Patterson et al., 2013). The model reproduced this suppressive effect (Figure 4B). This can be understood by considering the inferred probability of dependence and its influence on the divisive normalization. The probability inferred by the model increased with the similarity between adapter and test

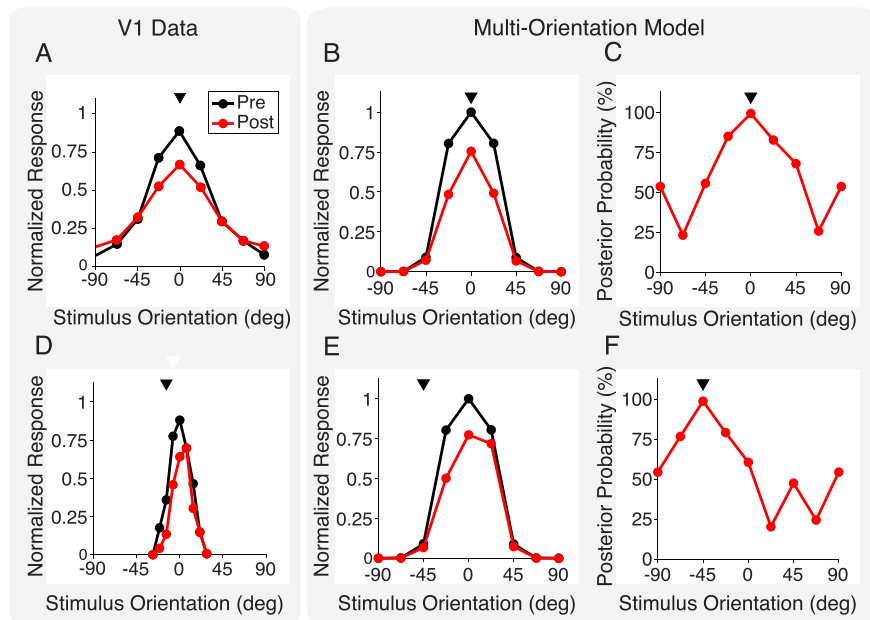


Figure 4. Effects of adaptation on neuronal response tuning curves. (A) Average orientation tuning curve responses in V1 data for pre (black line) and post (red line) adaptation. The neuron’s preferred orientation in the data (aligned to zero for visualization) was between 0° and 15° away from the adapter orientation (arrowhead). Adapted from Wissig and Kohn (2012). (B) Model prediction of neuronal response pre- and postadaptation, for a neuron that prefers a stimulus of 0° and was adapted to a stimulus of that orientation. (C) Inferred posterior probability, for a 0° normalization pool, that past and present stimuli are dependent, for the model neuron in (B). Note that the tuning curves of the inferred probability peak at the orientation of the adapter (arrowhead). (D) Same as (A) but for an adapter 14° away from preferred orientation, resulting in repulsion, namely, a shift of the tuning curve peak away from the adapter. Adapted from Müller et al. (1999). (E, F) Same as (B, C) but for an adapter 45° away from the model neuron’s preferred orientation. (B, E) Overall suppression strength in the model was controlled by a free parameter (Methods), which we set here to match the suppression level in (A) when both adapter and test have an orientation of 0° .

orientations (Figure 4C), resulting in large response suppression around the peak of the tuning curve and weak or no suppression toward the tails (Figure 4B).

We then considered the case in which the adapter orientation does not match the preferred orientation of the neuron. Experimentally, in this case, the tuning curve is both reduced in amplitude and shifted away from the adapter, an effect termed *repulsion* (Figure 4D; experimental data plotted from Müller et al., 1999; see also repulsion in other data across a range of timescales and adapting orientations on the flank in Dragoi et al., 2000, 2002; Felsen et al., 2002; Patterson et al., 2013; Wissig & Kohn, 2012). Note that the model is limited in predicting tuning curves to adapter orientations matched to one of four filter orientations detailed in the Methods, thus, the difference in adapter orientation between the plotted experimental data of Müller et al. (1999) and modeling predictions (14° in Figure 4D and 45° in Figure 4E). The repulsion could be explained qualitatively by our model (Figure 4E), because the dependence probability (and therefore the degree to which normalization was engaged) was determined by the match between test and adapter. To understand the model behavior, first recall that the model

uses four normalization pools each with a different orientation tuning. The adapter drives most strongly the normalization pool with matching preferred orientation (0° in Figure 4B; 45° in Figure 4E), regardless of the test stimulus. However, the normalization signal computed by such pools is used to normalize the neural response only to the degree that test and adapter are inferred dependent. Because the probability of dependence is highest when the adapting and test stimuli are matched in orientation (Figure 4F), adaptation is stronger when test stimuli are closer to the adapter. Our modeling framework thus provides a normative explanation from movie statistics for stimulus-specific adaptation (Solomon & Kohn, 2014).

Long-term adaptation and updating the model priors

We have focused so far on some classical adaptation experiments that examined how visual cortical responses to test stimuli are influenced by repeated exposure to an individual stimulus over a

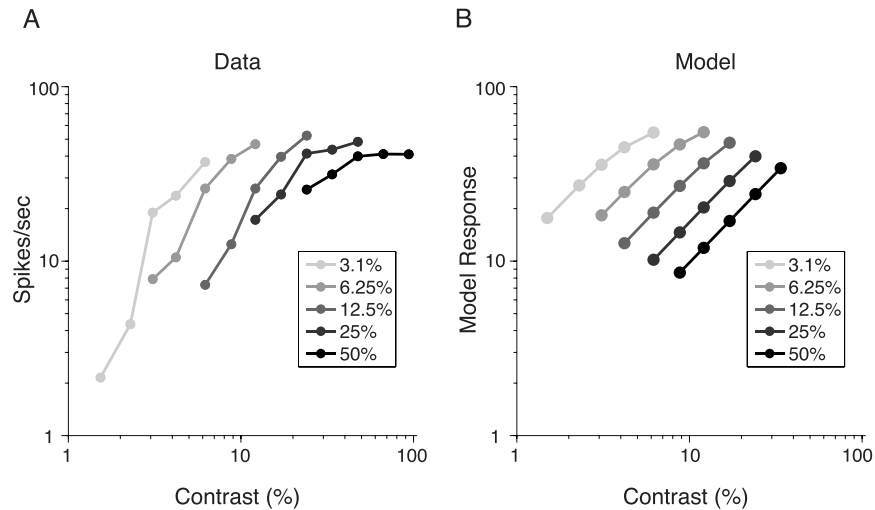


Figure 5. Contrast response functions after adaptation. (A) Neuronal response, in spikes per second, of a complex V1 neuron after adaptation to gratings of various contrasts for 80 s. Each line corresponds to a different grating contrast as indicated by the legend. After adaptation, the neurons were presented with test gratings of contrasts ranging from 1.5% to 94%. Adapted from Ohzawa et al. (1985). (B) Model predictions for the same paradigm as used to generate the experimental results in (A).

short period of time, on the scale of hundreds of milliseconds. In real life, we are constantly exposed to stimuli, and it is likely that the cortex is sensitive to different timescales through different mechanisms. Here we consider that the visual cortex keeps track of the changing environmental statistics by updating its prior knowledge about the environment. Specifically, we study how prolonged exposure to visual stimuli affects model responses by updating the prior probability of dependence. We first describe how we extend our model to capture longer-term adaptation and then apply it to shifts in contrast response curves with adaptation and to other recent data (Benucci et al., 2013). In the Discussion section, we also consider the implications of our model updating on capturing stronger changes in tuning curve suppression and repulsion with longer adaptation time.

In our model thus far, for any short sequence of eight adapting frames, the probability of dependence is obtained by combining the prior probability learned from natural movies with the evidence provided by the adapting and test frames (Figure 2A). This short-term model, which captures dependencies only over a small number of frames in the past, is unable to reproduce effects that require longer timescales of adaptation. We posit that on longer timescales of adaptation, the system is no longer just tuned to the statistics of natural movies but rather updates its model based on the current statistics of the experimental stimuli. We therefore extended the model and allowed it to update the prior probability of dependency as each new set of stimuli was presented (Figure 2B).

Contrast response following long-term adaptation

We next tested our model's ability to account for effects of contrast adaptation. To simulate the effects seen in contrast adaptation, we held the orientation of the grating fixed while varying the contrast of the test and adapter gratings. Figure 5A depicts the experimental findings for contrast adaptation that is typically measured on the order of seconds, namely, that increasing the contrast of the adapter shifts the contrast response curve down and to the right (Albrecht et al., 1984; Ohzawa et al., 1985), thereby reducing the contrast sensitivity of neurons.

To replicate this paradigm with our model, we repeated the adapting contrast for 30 frames, which for our movies approximates as 1 s. As each new stimulus was presented, the model updated its prior probability of dependency, based on the posterior inferred from the previous stimuli (Figure 2B). The model is able to generate similar qualitative results to the experimental data (Figure 5B) because of the role contrast plays in divisive normalization as well as the updating of the prior probability of dependency. As the adapting contrast increases, so too does the magnitude of the normalization signal (as expected by divisive normalization frameworks). This is due to a high probability of dependency between the adapting and test stimuli in our model following the adaptation. The repeated exposure of the model to the adapting contrast results in a prior probability heavily weighted toward the dependent regime, so a high-contrast past is inferred to be dependent with any contrast in the present. Note that the slope of the curve is different between the

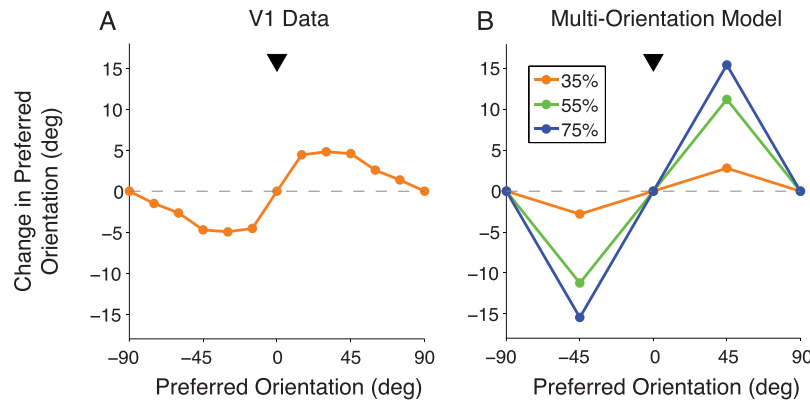


Figure 6. Effect of bias in the stimulus ensemble on shifts in tuning curve peak. (A) V1 neural responses after adaptation to stimuli that are shown more frequently at 0° (arrowhead; 35% bias) than at other orientations. Each point represents the average change in preferred orientation across all neurons with preadaptation preference indicated in the abscissa. Data adapted from Benucci et al. (2013). (B) Model predictions for three different frequency biases (indicated in the inset).

model and neuron because of the choice of prior in the GSM model (Rayleigh for computational tractability), which influences the exponent of the normalization term and therefore the slope of the contrast response function.

Without this updating in the model, when there is a significant difference (at least fourfold) between the adapter and test contrasts, there is a decreased predicted probability of dependence between the past and the present. This can lead to a situation in which a moderate contrast adapter may facilitate a low-contrast test on short timescales in our model. We return to this in the Discussion section.

Population-level response equalization with longer-term adaptation

We next considered some recent data (Benucci et al., 2013) in which a longer-term adaptation paradigm has been used experimentally to test adaptation to a distribution of environmental statistics, rather than to a single stimulus. Benucci et al. (2013) considered an ensemble of gratings of different orientations, presented over several seconds. Each orientation was presented at random times with equal frequency except for one orientation (the biased stimulus orientation), which was shown with a greater frequency. To measure the effects on V1 populations, Benucci et al. (2013) measured the tuning curves for each group of neurons sharing a specific orientation preference. When the stimuli were drawn from a distribution biased to a single orientation, the adapted responses still reproduced the classical effects of tuning curve suppression and repulsion (Figure 6A).

We first verified that for the biased stimuli, as for the classical adaptation paradigm (which may be seen as an extreme 100% bias case, because a single orientation is

presented 100% of the time and all other orientations are not presented at all), our model learns to update the prior probability of dependence over time, leading to an increased probability for the normalization group whose orientation preference matches the biased orientation in the stimulus ensemble. We next confirmed that our model achieves both repulsion and suppression with this paradigm. In addition, the model is able to predict how the strength of the repulsive shift is affected by the degree of bias to a single orientation (Figure 6B). This is a prediction of the model that has yet to be tested experimentally.

The experimental paradigm described above offers a richer set of data beyond the classical suppression and repulsion. In this paradigm, a stimulus of a specific orientation (the adaptive stimulus) is shown with a greater frequency than other orientations. One of the main experimental findings of Benucci et al. (2013) was that the effects of adaptation on tuning curves compensate exactly for the overrepresented stimulus. The time average of population responses to stimuli drawn from the distribution biased to a single orientation was not significantly different from that of stimuli drawn from the uniform distribution, a phenomenon that has been termed *equalization* (Benucci et al., 2013; Figure 7G). Stated differently, on one hand, the biased orientation was presented more frequently than the other orientations, but on the other hand, the response of neurons preferring the biased orientation was suppressed more strongly than neurons preferring other orientations. These two effects counterbalanced each other, such that the average response over the entire stimulus ensemble did not differ between neurons with different orientation preferences.

To simulate the equalization experiment, we created grating stimulus ensembles drawn from a biased distribution (Figure 7A). We then exposed the model to both the biased distribution and the uniform distribu-

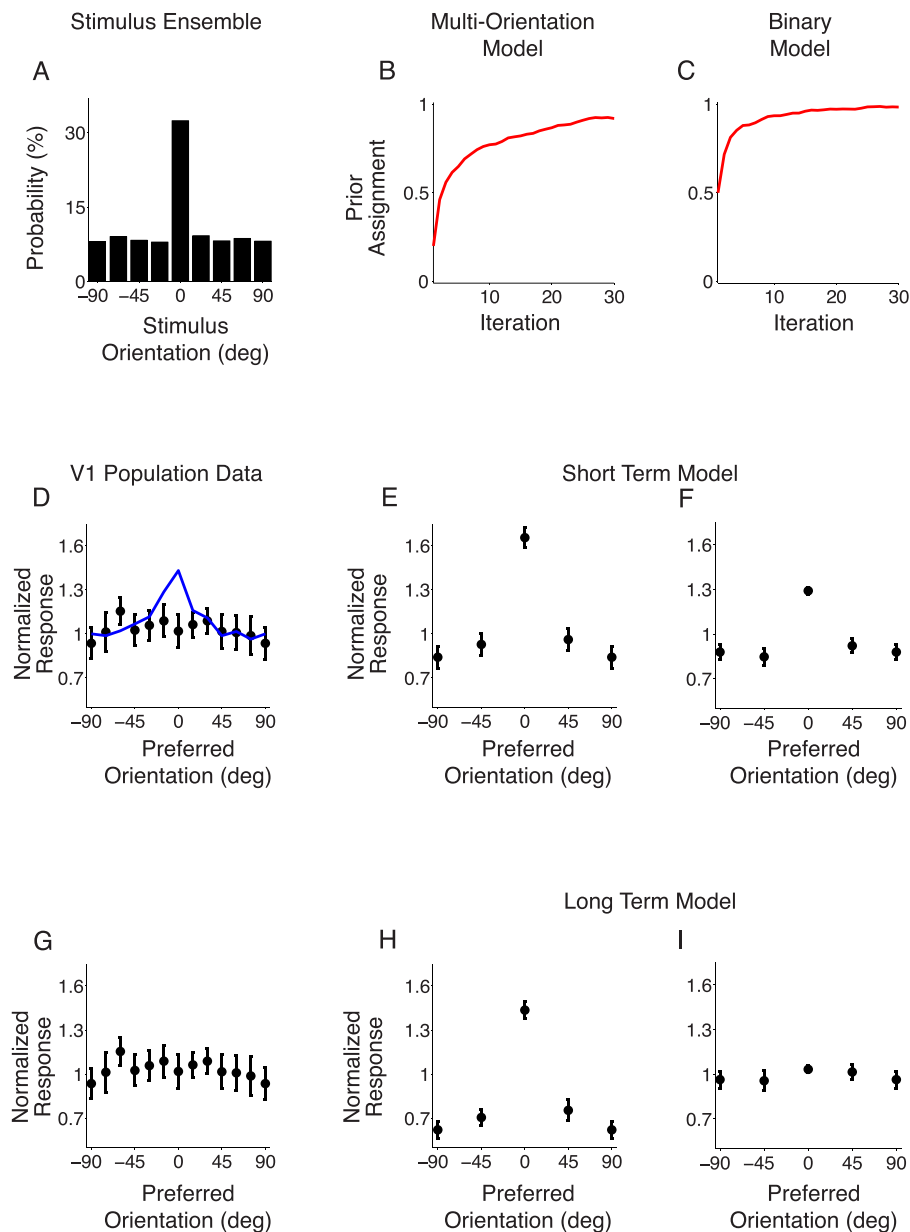


Figure 7. Neuronal response equalization. (A) The probability of presentation for different stimulus orientations. Zero degrees is shown 35% of the time; the others are shown less than 10% of the time. (B, C) The prior probability of dependence for the biased orientation as a function of iteration number for the two models. (D) Experimental finding showing V1 population responses to gratings drawn from the biased distribution, normalized by the mean responses to gratings drawn from the uniform distribution. Error bars represent ± 1 SD. The blue line is a prediction of normalized response from a model without adaptation. Adapted from Benucci et al. (2013). (E, F) Population responses of the short-term model to the biased stimulus ensemble, normalized to the uniform ensemble as in (D). We simulated one experiment by presenting the stimulus for 300 frames (roughly 10 s), averaging the responses and normalizing by the uniform distribution. We then repeated the experiment six times; the error bars represent ± 1 SD across repeated experiments, similar to (D). (G) V1 population responses replotted from (D). (H, I) Same as (E, F) but for the long-term model outlined in Figure 2B. (E, F, H, I) Overall suppression strength in the model was controlled by a free parameter (Methods), which we set here to match the suppression level in Benucci et al. (2013) when both adapter and test have an orientation of 0° .

tion. As expected from Benucci et al. (2013; Figure 7D, blue curve), using the short-term model to predict equalization failed to adequately suppress the biased orientation. As a result, the response of neurons preferring the biased orientation was much larger than

neurons with different preferences, as illustrated by the central peak in Figure 7E.

We therefore asked whether the long-term model could account for population-level equalization. We started from the parameters learned initially on the

natural movie ensemble (approximately a flat prior probability of dependence across orientations) and updated the prior as the biased stimulus ensemble was presented. As shown in Figure 7B, the prior correctly increased for the biased orientation. The prior for the orthogonal orientation decreased, whereas it did not change as much for the intermediate orientations (not shown). This updated prior led to stronger suppression overall (compare Figure 7H to the short-term model, Figure 7E). Nonetheless, the model was unable to equalize population responses, as illustrated by the central peak in Figure 7H.

To understand this failure of the model, we must first note that the findings in Benucci et al. (2013) have been described as arising from a combination of stimulus-specific and neuronal-specific adaptation. First, the responses to test stimuli that are matched to the biased orientation are more strongly suppressed than to stimuli with different orientations; this effect is consistent with sensitivity to the statistical similarity of the past and present inputs. Furthermore, regardless of the test stimulus, the responses of neurons whose preference matches the biased orientation are more strongly suppressed than those of other neurons; this effect has been ascribed to neuronal fatigue (i.e., a decrease in responsivity for neurons that are stimulated more often or more strongly).

Equalization via neuronal-specific normalization model

The model we have considered thus far, which we have denoted the multiorientation model, lacks the second component of neuronal-specific adaptation and thus cannot capture equalization. More specifically, the model uses four normalization pools, each comprising a set of past RFs with a given orientation (Figure 8A, left). When a short movie sequence is presented, the model computes the probability of dependence between the present RFs and the past RFs of each of the normalization pools. Subsequently, the response of the model neuron is normalized by each of these pools according to the estimated probabilities. The resulting normalization signal depends only on the properties of the stimulus and is identical for all RFs in the present, regardless of their preferred orientation (Figure 8B, left). For example, if both the adapter and test stimuli are vertical, then both a neuron with preferred vertical orientation and a neuron with preferred horizontal orientation will be suppressed divisively by the vertical normalization pool (Figure 8A, left). As a consequence, even with the biased stimulus ensemble, model neurons tuned to the biased orientation are not suppressed more strongly than other neurons, and population responses cannot be equalized (Figure 7H).

We therefore considered a different choice of divisive normalization pools, which we denote the binary model (see the Methods section for details). In this version, each model neuron has a private normalization pool with an orientation that is matched to its preferred orientation (Figure 8A, right). The divisive normalization signal for each neuron is still weighted by the statistical similarity between past and present stimuli, but it is tuned only to the neuron's preferred orientation. Therefore, each neuron in this model has a different tuning of the normalization signal (Figure 8B, right). For example, if both the adapter and test stimuli are vertical, this will result in strong suppression for a vertical neuron but not for a horizontal neuron (Figure 8A, right). This binary model implements a neuronal-specific normalization.

We then tested whether the binary model with the biased stimulus ensemble could account for population-level equalization. We started from the prior probability of dependence learned initially on the natural movie ensemble. In this case, each neuron has its own private normalization pool. First, we computed the averaged responses to the biased stimulus ensemble for the short-term model (i.e., without updating the prior). Figure 7F shows that, in this model, the difference in the average responses across neurons was smaller than in the multiorientation model (Figure 7E). This is because the neuron whose preference matched the biased orientation was driven more strongly by the stimuli, but also suppressed more strongly by the context, as explained above. However, the population responses were not precisely equalized.

We then considered the long-term version of the binary model. When we updated the prior to reflect the biased ensemble, the prior correctly increased for the neuron whose preference matched the biased orientation (Figure 7C). As a result of the prior updating, suppression increased for the neuron whose preference was matched to the biased orientation more than for the other neurons (not shown); this relative increase in suppression reflected the stimulus ensemble statistics and led to equalization of population responses (illustrated by the lack of a central peak in Figure 7I).

To further demonstrate that equalization was due to the correct updating of the prior, we also considered stimulus ensembles with different amounts of bias. We found that ensembles with larger bias led to higher values for the updated prior of dependence reflecting precisely the ensemble bias level (Supplementary Figure S1D, E); this in turn led to stronger relative suppression for the matched neuron, and therefore equalization was maintained for a range of bias levels (Supplementary Figure S1A, B). However, when the bias became too large, the updated prior saturated (i.e., it reached a value of 1; Supplementary Figure S1F), and therefore the relative suppression could not be further increased,

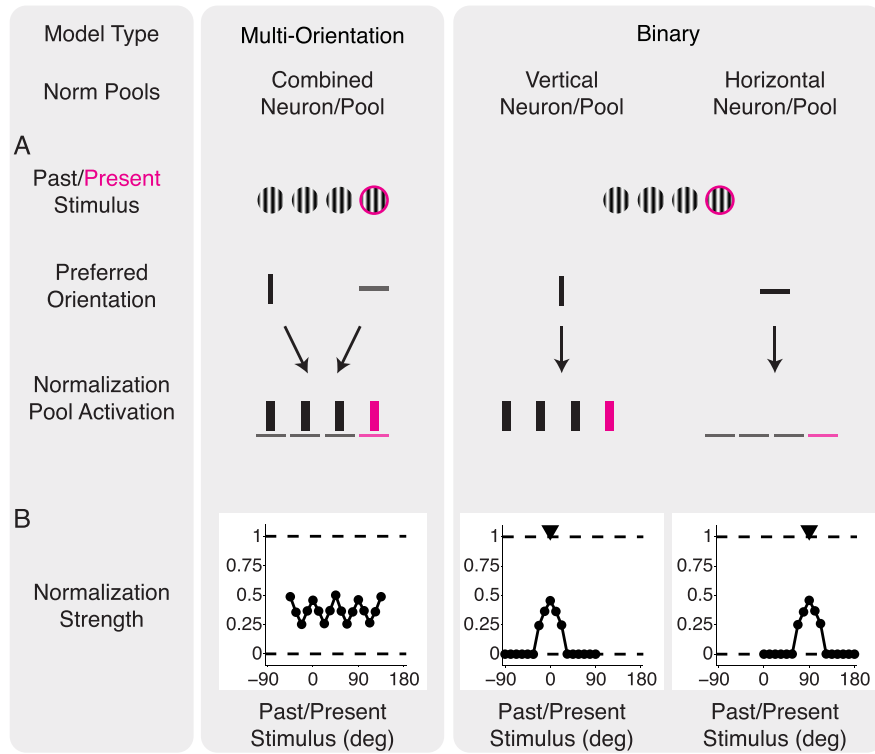


Figure 8. Comparison of normalization signals in the binary model and multiorientation model. We include only two filter orientations for illustration purposes, although the models use four orientations as detailed in the Methods. (A) Both models were exposed to a single set of past (black) and present (magenta) vertical stimuli (top row). We consider, for each model, two model neurons with vertical and horizontal preferred orientation (middle row). Bottom row: Two normalization pools are illustrated, with different orientation preferences. Bar thickness represents a cartoon of the strength of the output of the filters in each normalization pool. In the multiorientation model (left), both neurons are normalized by the most active normalization pool. In the binary model (right), each neuron has its private normalization pool, and therefore the normalization signal is stronger for the vertical neuron. (B) Normalization strength, which is the amount of suppression of the postadaptation response relative to the preadaptation response, $1 - \frac{E[g_t|x_p, x_t]}{E[g_t|x_t]}$, for each normalization pool at a wide range of past/present stimuli orientations. The multiorientation model (left) shows consistently high normalization strength across a wide range of orientations. In the binary model (right), only the ranges around the preferred orientations (arrowheads) show substantial normalization. The unevenness in the multiorientation model curve is due to using only four neuronal orientations, as detailed in the Methods.

leading to a breakdown of equalization (Supplementary Figure S1C) similar to the experimental observation of Benucci et al. (2013). Changes in contrast follow the same response trend as changes in bias. Increasing contrast generally leads to larger probability of dependence and therefore stronger suppression, as we have seen in the spatial domain (Coen-Cagli et al., 2012).

Although the binary model with an updating prior led to population-level equalization, we found that it could not account for tuning curve repulsion (Supplementary Figure S2E). This is because the binary model captures neuronal-specific but not stimulus-specific adaptation, because of the choice of the normalization pool and how the statistical similarity between adapter and target is computed. In particular, for the binary model, the inferred probability of dependence is determined by how similar both the adapter and test stimulus are to the neuron's preferred

orientation. For example, in Supplementary Figure S2E and S2F, we considered a neuron tuned for vertical orientation (0°) and used an adapter fixed at 45° ; in this case, regardless of the test stimulus, the inferred probability was close to 0, because the adapter was largely different from the neuron's preference (Supplementary Figure S2F), and therefore test stimuli at $+45^\circ$ or -45° did not lead to different responses (Supplementary Figure S2E).

In summary, our modeling framework explained both short-term adaptation effects on tuning curves and longer-term adaptation effects on contrast and population responses, as sensitivity of the visual cortex to the statistics of the visual environment on different timescales. However, each set of effects could be captured only by a particular implementation of the model (and a respective normalization signal that is either stimulus specific or neuronal specific), an issue that we further address in the Discussion.

Discussion

We have considered the link between movie statistics, divisive normalization models, and adaptation phenomena in primary visual cortex. Our approach was to learn the parameters of an interpretive model of adaptation based on scene statistics, an approach more prevalently used in the spatial context domain (Coen-Cagli et al., 2012; Rao & Ballard, 1999; Schwartz & Simoncelli, 2001; Spratling, 2012). Although descriptive models address what neural circuits do and mechanistic models describe how they do it, interpretive models focus on why the neural systems operate in the way they do (Dayan & Abbott, 2001). Descriptive and mechanistic models typically rely on data or circuitry, respectively, as the basis for their construction. Interpretive models (which may be built on top of descriptive or mechanistic approaches) use computational principles, such as efficient coding or Bayesian inference, as their foundation. Here we assume that when confronted with sensory stimuli over time, the brain attempts to reduce the dependency in the visual inputs. This is implemented in our model by performing what amounts to Bayesian inference based on its prior belief about the world and the likelihood for the current stimulus. In our modeling framework, inference amounts to a generalized form of divisive normalization.

Here we have focused on the implications of this scene statistics approach for adaptation. Many models of adaptation are descriptive in nature; they describe observed effects using a parameterized function that can be fit to data. Descriptive models of adaptation have included divisive normalization, a ubiquitous computation in neural systems (Carandini & Heeger, 2012; Dhruv, Tailby, Sokol, & Lennie, 2011; Heeger, 1992). There have also been other models of adaptation that are more mechanistic in nature (Cortes et al., 2012; Teich & Qian, 2003). Our model does not challenge descriptive or mechanistic models of temporal contextual effects but rather incorporates some of their aspects, such as divisive normalization, and provides a principled explanation for why cortex performs such computation.

We have shown that our modeling approach can explain some classical adaptation effects as well as a more recent phenomenon of equalization. However, clearly the approach has limitations and does not capture the full set of phenomena for adaptation. First, the model in its present form does not include surround influences and cannot capture disinhibition of the surround nor interesting data on facilitation and attractive shifts of tuning curves (Solomon & Kohn, 2014; Webb et al., 2005; Wissig & Kohn, 2012). It would be interesting to consider extensions of the model to capture both spatial (Coen-Cagli et al., 2012)

and temporal contextual influences. Second, although the modeling approach offers a step forward in updating scene statistics normalization frameworks and addressing timescales of adaptation, the model can account for only some qualitative rather than quantitative influences of time. It does not capture more quantitative aspects of adaptation decay over time and recovery, as some models of synaptic depression (e.g., Chance, Nelson, & Abbott, 1998) or power law scaling (Drew & Abbott, 2006). As we discuss below in the Timescales of Adaptation section, the updating of the prior can reach a ceiling, and adaptation at longer timescales (and their progression over time) may be better explained by updating of the covariance matrix in our framework. In addition, as noted in the Results section, for brief adaptation to contrast in our short-term model, adaptation to a high contrast that is followed by a very low contrast may result in facilitation. We are not aware of experimental data that support this prediction. But contrast adaptation acts differently at different timescales of onset and steady state (Bonds, 1991; Müller, Metha, Krauskopf, & Lennie, 2001). Here we have focused on the most basic contrast adaptation effect observed across many labs, typically in long-term adaptation experiments, and shown that with updating of the model prior, we observe results consistent with the data.

We addressed two important issues that have received less attention in the literature on scene statistics and adaptation: (a) stimulus-specific versus neuronal-specific contextual modulation and (b) learning on different timescales. By focusing on a broader set of data, we both highlight some of the complexities of formulating a more comprehensive theory of adaptation and provide a step forward in this direction. We next discuss each of these issues in turn and then directions on how to unify these concepts into a more complete interpretive model of cortical adaptation. We also compare our approach to other adaptation models and scene statistics approaches in the literature.

Stimulus-specific versus neuronal-specific contextual modulation

Our modeling approach distinguishes between stimulus-specific and neuronal-specific effects, a distinction already implied in the literature (Benucci et al., 2013; Solomon & Kohn, 2014). Adaptation is often described as stimulus specific, that is, the strongest effects occur when the features of the adapter and the test stimulus match, which, among other effects, results in repulsive shifts of tuning curves (e.g., Dragoi et al., 2002; Felsen et al., 2002; Müller et al., 1999; Patterson et al., 2013). In our modeling, we have captured this set of phenomena with an implementation that assumes the

strength of the normalization signal is determined by the statistical similarity between adapter and test stimuli (Figure 8, left).

There have also been suggestions of neuronal-specific effects. For instance, fatigue could be induced by prolonged stimulation of a neuron and hence depend on how well the adapter is matched to the neuron's RF (Albrecht et al., 1984; Carandini & Ferster, 1997; Crowder et al., 2006; Giaschi, Douglas, Marlin, & Cynader, 1993; Movshon & Lennie, 1979; Ohzawa et al., 1982). Population equalization data (Benucci et al., 2013) provided a strong test case, which, in our modeling, could be captured only by assuming that the normalization strength is determined by the statistical similarity of the stimuli as viewed through each neuron's RF (Figure 8, right). Our results show that neuronal-specific adaptation is sufficient to explain population-level equalization, but a stimulus-specific component is also necessary, to account for repulsive shifts of tuning curves, as also suggested by Benucci et al. (2013).

Standard divisive normalization models of adaptation have typically incorporated neuronal-specific response reduction (Carandini, Heeger, & Senn, 2002; Heeger, Simoncelli, & Movshon, 1996; Heeger, 1992; Sinz & Bethge, 2013), although some interpretive models of adaptation are perhaps stimulus specific (Lochmann et al., 2012; Stevenson et al., 2010; Wainwright et al., 2002). Note that many models do not make such distinctions explicit and mostly do not study a range of data that require or strongly distinguish between both forms of model. By studying a more comprehensive range of adaptation data, we have had to more deeply examine what aspects can or cannot be explained by current versions of divisive normalization models motivated by scene statistics. The modeling framework presented here explains both neuron- and stimulus-specific adaptation effects as determined by the statistical similarity of the past and present inputs but using two different metrics of similarity.

Specifically, we considered two classes of divisive normalization model. The first model, which we termed the *multiorientation model*, amounted to stimulus-specific adaptation. The model learned a common normalization signal for all oriented RFs, regardless of their orientation preference, which effectively depended only on the similarity between the adapting and test stimuli. This model mimicked the previous approach we used to capture spatial context neurophysiology data (Coen-Cagli et al., 2012) and reproduced classical short-term effects such as response suppression and tuning curve repulsion (Figure 4), as well as contrast adaptation (Figure 5). However, the stimulus-specific model could not capture response equalization (Benucci et al., 2013). We therefore implemented a binary model with independent normalization pools, which amounted to neuronal-specific normalization, and showed that this

could capture equalization (Figure 7). However, response suppression was purely a measure of how well the adapter stimulus drove the RF of the neuron, and as a result, this model was unable to replicate tuning curve repulsion (Supplementary Figure S2).

This dichotomy of stimulus specific versus neuronal specific may reflect a difference in underlying mechanisms for contextual effects. Neural fatigue mechanisms, such as hyperpolarization (Carandini & Ferster, 1997; Sanchez-Vives, Nowak, & McCormick, 2000), have been purported to drive response reduction. The mechanism underlying repulsive shifts has for the most part been laid at the feet of intracortical connections, such as synaptic depression (Abbott, Varela, Sen, & Nelson, 1997; Adorján, Piepenbrock, & Obermayer, 1999; Carandini et al., 2002; Chance et al., 1998; Chung, Li, & Nelson, 2002; Finlayson & Cynader, 1995; Nelson, 1991). However, neither of these mechanisms explains both effects simultaneously (for review, see Solomon & Kohn, 2014). As expected, there is probably no single unifying mechanistic model of adaptation, but rather multiple mechanisms that are combined in different ways to reproduce the observed effects (Grill-Spector, Henson, & Martin, 2006; Ursino, Magosso, & Cuppini, 2008).

These issues are also of general importance to contextual phenomena beyond adaptation. There have been previous indications that in the spatial domain, both types of modulation might be important (Trott & Born, 2015). For instance, some neurophysiology surround data are well captured by a suppressive term determined by the orientation match between the surround and test stimulus (Sillito, Grieve, Jones, Cudeiro, & Davis, 1995), thus relating to stimulus-specific suppression. Our spatial context model for capturing surround suppression (Coen-Cagli et al., 2012) was also based on a stimulus-specific formulation. On the other hand, the tilt illusion due to surrounding stimuli (and, similarly, tilt after effect in the temporal domain) has typically been modeled by neuronal-specific suppression (Clifford et al., 2000; Schwartz et al., 2009; Serès et al., 2009).

This distinction highlights the importance of working out in the future a proper interpretive model that can encompass the full wealth of data, for instance, by adding another layer hierarchically on top of the two models, that infers from the statistics of the inputs the probability of the stimulus-specific versus neuronal-specific model components, thereby amounting to a mixture of these components.

Timescales of adaptation

We have developed an adaptation model from natural movie statistics, which captures multiple

timescales. In early visual cortex, adaptation is known to act on a broad range of timescales, ranging from hundreds of milliseconds to seconds (Müller et al., 1999; Ohzawa et al., 1985), seconds to minutes (Movshon & Lennie, 1979; Sharpee et al., 2006), and even hours to days (Vul, Krizay, & MacLeod, 2008; Wolfe & O’Connell, 1986), with both quantitative and qualitative differences (Kohn, 2007; Solomon & Kohn, 2014). The main new aspect of our work from this perspective is that we have developed a principled modeling framework that explains adaptation as a form of divisive normalization optimized to the statistics of natural movies, while also providing a way to update the models when the environmental statistics change. We have applied the model to adaptation phenomena ranging from hundreds of milliseconds to seconds, providing a framework for adaptation effects at multiple timescales.

Specifically, we started from the hypothesis that the responses of V1 neurons represent Bayesian inference about short sequences of visual inputs and that such inference leads to divisive normalization only when a statistical dependence between temporally adjacent inputs is detected (Figure 1). Following Bayesian principles, the judgment about statistical dependence relies on two factors: the prior probability that visual inputs are dependent (e.g., based on knowledge derived from natural movies) and the likelihood that a given sequence of inputs (e.g., the particular experimental stimulus sequence) is indeed statistically dependent. The priors are initially learned from an ensemble of natural movies and then held fixed when we capture short-term adaptation data. The prior parameterizes how often visual sequences encountered in natural movies are dependent (or independent). The short-term adaptation data are interpreted in light of the learned priors. In this framework, short-term adaptation effects are captured by the divisive normalization operation that the model performs on any short sequence of inputs (Figure 2A).

However, this approach does not capture adaptation effects that occur on longer timescales. First, longer-term adaptation can result in shifts in contrast response functions and in neuronal population response equalization. Second, adaptation effects, such as suppression and repulsion of tuning curves, are strengthened by longer-term adaptation (Dragoi et al., 2000; Patterson et al., 2013).

We therefore considered how to capture longer timescales of adaptation within our framework. One approach could be to lengthen the time window over which we learn statistical dependencies, for instance, from hundreds of milliseconds to seconds or minutes. However, learning the parameters of our model over long timescales is computationally intractable, increasing in complexity with the increasing timescale.

Furthermore, as the time window is expanded to the order of seconds, the model would potentially miss effects occurring on the order of tens to hundreds of milliseconds.

We therefore took an alternative approach of maintaining a short temporal length of the input window but updating the model parameters to track the statistics of the environment (Figure 2B). We thus viewed the learned parameters as dynamic variables, representative of the changing visual environment. We specifically considered the updating of the prior and showed that it allows the model to capture shifts in contrast response functions (Figure 5) and the equalization of population responses (Figure 7).

However, our model contains two learned parameters that control the divisive normalization, namely, the prior and the covariance matrices. Similar to the method that we used to update the prior, the covariance in our framework could also be updated over time to reflect the changing statistics of the visual environment. Such covariance matrices capture long-term knowledge about how dependent and independent input sequences appear, which could be represented, for example, in the synaptic weights of recurrent connections (Coen-Cagli et al., 2012). We can approximate the effect of such changes by directly scaling the learned covariances (Figure 9). We demonstrate as a proof of concept that this can result in increased suppression and repulsion with larger durations of adaptation for stimuli confined to the classical RF of the neuron (Figure 9), as shown in experimental literature (Dragoi et al., 2000; Patterson et al., 2013). Although updating the prior can also result in increased suppression and repulsion if one starts from a low prior probability of dependence, we found that from our starting conditions of the parameters learned from natural movies, the probability of dependence between like orientations quickly saturates and prohibits the type of increased changes in tuning curves that are seen experimentally over longer timescales.

Our modeling is a step toward showing how normalization models from scene statistics may be adjusted to account for adaptation at multiple timescales. Future work would involve combining the short-term estimation of neural responses with the long-term updating of both the prior and the covariances. However, our model is currently limited in the range of timescales effects we can account for. Although we have explained how our model may account for some duration scaling effects of adaptation, the model would need to be extended (e.g., to include surrounding filters and disinhibition) to account for violations of duration scaling found in neurophysiology V1 data (Patterson et al., 2013; Solomon & Kohn, 2014).

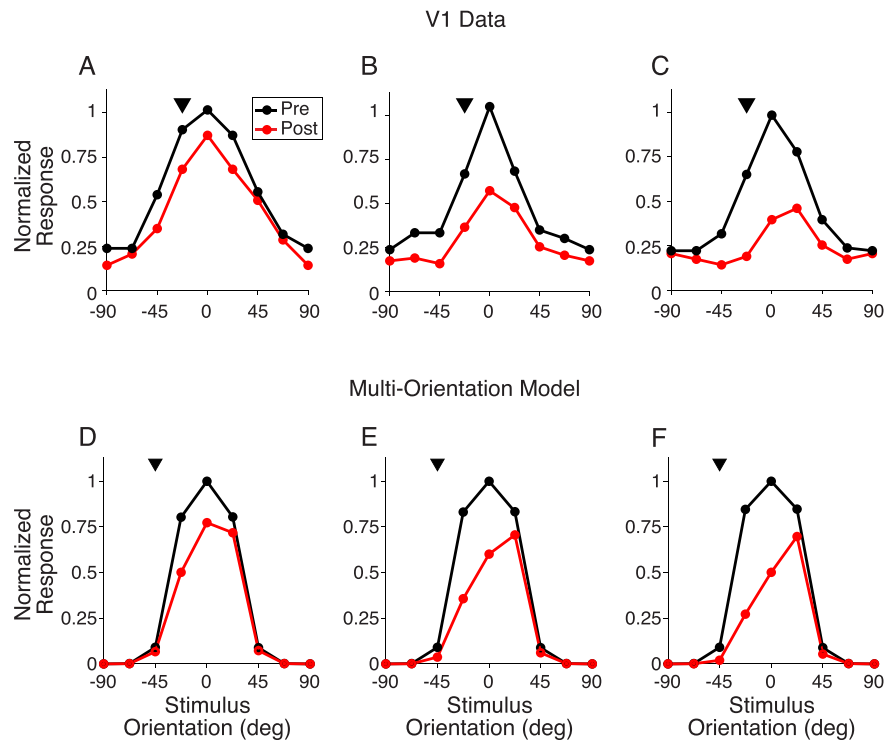


Figure 9. Tuning curve repulsion strength for increasing adaptation times. (A) Orientation tuning curve responses in V1 data for pre (black line) and post (red line) adaptation. The neuron's preferred orientation in the data (aligned to zero for visualization) was 45° away from the adapter orientation (arrowhead). The neuron was adapted to the gratings for 400 ms. Adapted from Patterson et al. (2013). (B, C) Same as (A) but for adapting periods of 4 and 40 s, respectively. (D) Model prediction of neuronal response pre and post adaptation, for a neuron that prefers a stimulus of 0° and was adapted to a stimulus 45° away. (E, F) Same as (D) but for a scaling of the learned covariances by a factor of 2 and 4, respectively.

There is also interest in addressing longer-term perceptual phenomena. Some studies have shown increased perceptual effects with adaptation time (e.g., Greenlee, Georgeson, Magnussen, & Harris, 1991; Greenlee & Magnussen, 1987). Perceptually, there have also been intriguing adaptation effects, linked to early visual cortex, which can even last orders of magnitude longer, on the scale of hours to days (Vul et al., 2008; Wolfe & O'Connell, 1986). Furthermore, recent work suggests that the cortex has a memory of its distant past, and even after interruption, which seemingly negates the effects of adaptation, adaptation reemerges (Bao & Engel, 2012; Chopin & Mamassian, 2012).

Relation to other models of adaptation

From the scene statistics perspective, it is well documented that natural movies contain temporal regularities (Dong & Atick, 1995; Hyvärinen, Hurri, & Hoyer, 2009; Kayser, Einhäuser, & König, 2003; Schwartz et al., 2007; Simoncelli & Olshausen, 2001). Temporal statistics have been a basis for some previous adaptation models. Redundancy reduction frameworks have previously been considered for temporal context

effects, explaining adaptation as a form of decorrelation (Barlow & Földiák, 1989).

More closely related to our work, Wainwright et al. (2002) suggested that redundancy reduction can be achieved by a form of divisive normalization optimized to natural scene statistics and so captured effects of tuning curve changes and shifts in contrast response curves (Wainwright et al., 2002), in what may be considered a long-term adaptation model for all of the simulations. This model was trained on still images rather than temporal structure in movies. Unlike our model in which the parameters are updated over time as new stimuli are presented, this model relearned its weights by mixing stimuli from the old visual environment (e.g., natural images) with the new visual environment (e.g., gratings).

Previous predictive models of adaptation share some aspects, but not others, with our model. For instance, the models of Lochmann et al. (2012), Stevenson et al. (2010), and Stocker and Simoncelli (2006) were based on Bayesian inference but not trained on natural stimuli. Others were trained on natural movies with a single pixel at each frame to capture gain control in the fly retina (Buiatti & van Vreeswijk, 2003). Still other models have applied Bayesian frameworks for learning

the temporal structure of multiple scales in natural sounds but without linking to neural adaptation phenomena (Turner & Sahani, 2008).

Keywords: adaptation, Bayesian, scene statistics, divisive normalization

Acknowledgments

This work was supported by the National Institutes of Health (NIH) Grant CRCNS-EY021371 and the ARMY Research Office Grant 58760LS. M. S. was supported in part by the NIH Medical Scientist Training Program training Grant T32-GM007288. We are very grateful to A. Kohn, C. Henry, and J. L. Pena for discussion and comments on the manuscript.

Commercial relationships: None.

Corresponding author: Michael Snow.

Email: michael.snow@med.einstein.yu.edu.

Address: Department of Systems & Computational Biology, Albert Einstein College of Medicine, Bronx, NY, USA.

References

- Abbott, L. F., Varela, J. A., Sen, K., & Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, *275*, 220–224.
- Adorján, P., Piepenbrock, C., & Obermayer, K. (1999). Contrast adaptation and infomax in visual cortical neurons. *Reviews in the Neurosciences*, *10*, 181–200.
- Albrecht, D. G., Farrar, S. B., & Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology*, *347*, 713–739.
- Albrecht, D. G., & Geisler, W. S. (1991). Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, *7*, 531–546.
- Andrews, D. F., & Mallows, C. L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society. Series B*, *36*, 99–102.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, *61*, 183–193.
- Bao, M., & Engel, S. A. (2012). Distinct mechanism for long-term contrast adaptation. *Proceedings of the National Academy of Sciences, USA*, *109*, 5898–5903.
- Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (Vol. 1, pp. 217–234). Cambridge, MA: MIT Press.
- Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences*, *24*, 602–607.
- Barlow, H., & Földiák, P. (1989). Adaptation and decorrelation in the cortex. In R. Durbin, C. Miall, & G. Mitchison (Eds.), *The computing neuron* (pp. 54–72). Reading, MA: Addison-Wesley.
- Benucci, A., Saleem, A. B., & Carandini, M. (2013). Adaptation maintains population homeostasis in primary visual cortex. *Nature Neuroscience*, *16*, 724–729.
- Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*, 83–87.
- Bonds, A. B. (1991). Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neuroscience*, *6*, 239–255.
- Buiatti, M., & van Vreeswijk, C. (2003). Variance normalisation: A key mechanism for temporal adaptation in natural vision? *Vision Research*, *43*, 1895–1906.
- Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., & Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience*, *25*, 10577–10597.
- Carandini, M., & Ferster, D. (1997). A tonic hyperpolarization underlying contrast adaptation in cat visual cortex. *Science*, *276*, 949–952.
- Carandini, M., & Heeger, D. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.
- Carandini, M., Heeger, D., & Senn, W. (2002). A synaptic explanation of suppression in visual cortex. *Journal of Neuroscience*, *22*, 10053–10065.
- Carandini, M., Movshon, J. A., & Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology*, *37*, 501–511.
- Chance, F. S., Nelson, S. B., & Abbott, L. F. (1998). Synaptic depression and the temporal response characteristics of V1 cells. *Journal of Neuroscience*, *18*, 4785–4799.
- Chopin, A., & Mamassian, P. (2012). Predictive

- properties of visual adaptation. *Current Biology*, *22*, 622–626.
- Chung, S., Li, X., & Nelson, S. B. (2002). Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses in vivo. *Neuron*, *34*, 437–446.
- Clifford, C. W., & Rhodes, G. (2005). *Fitting the mind to the world: Adaptation and after-effects in high-level vision. Advances in visual cognition*. Oxford, UK: Oxford University Press.
- Clifford, C. W., Webster, M. A., Stanley, G. B., Stocker, A. A., Kohn, A., Sharpee, T. O., & Schwartz, O. (2007). Visual adaptation: Neural, psychological and computational aspects. *Vision Research*, *47*, 3125–3131.
- Clifford, C. W., Wenderoth, P., & Spehar, B. (2000). A functional angle on some after-effects in cortical vision. *Proceedings. Biological Sciences/The Royal Society*, *267*, 1705–1710.
- Coen-Cagli, R., Dayan, P., & Schwartz, O. (2009). Statistical models of linear and nonlinear contextual interactions in early visual processing. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Nips* (pp. 369–377). Cambridge, MA: MIT Press.
- Coen-Cagli, R., Dayan, P., & Schwartz, O. (2012). Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Computational Biology*, *8*, e1002405.
- Coen-Cagli, R., & Schwartz, O. (2013). The impact on midlevel vision of statistically optimal divisive normalization in V1. *Journal of Vision*, *13*(8):13, 1–20, doi:10.1167/13.8.13. [PubMed] [Article]
- Cortés, J. M., Marinazzo, D., Seriès, P., Oram, M. W., Sejnowski, T. J., & van Rossum, M. C. W. (2012). The effect of neural adaptation on population coding accuracy. *Journal of Computational Neuroscience*, *32*, 387–402.
- Crowder, N. A., Price, N. S. C., Hietanen, M. A., Dreher, B., Clifford, C. W., & Ibbotson, M. R. (2006). Relationship between contrast adaptation and orientation tuning in V1 and V2 of cat visual cortex. *Journal of Neurophysiology*, *95*, 271–283.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience* (Vol. 806). Cambridge, MA: MIT Press.
- Dayan, P., Sahani, M., & Deback, G. (2002). Adaptation and unsupervised learning. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems 15* (pp. 237–244). Cambridge, MA: MIT Press.
- Dhruv, N. T., Tailby, C., Sokol, S. H., & Lennie, P. (2011). Multiple adaptable mechanisms early in the primate visual pathway. *Journal of Neuroscience*, *31*, 15016–15025.
- Dong, D., & Atick, J. J. (1995). Statistics of natural time-varying images. *Network: Computation in Neural Systems*, *6*, 345–358.
- Dragoi, V., Sharma, J., Miller, E. K., & Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature Neuroscience*, *5*, 883–891.
- Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron*, *28*, 287–298.
- Drew, P. J., & Abbott, L. F. (2006). Models and properties of power-law adaptation in neural systems. *Journal of Neurophysiology*, *96*, 826–833.
- Fairhall, A. L., Lewen, G. D., Bialek, W., & de Ruyter Van Steveninck, R. R. (2001). Efficiency and ambiguity in an adaptive neural code. *Nature*, *412*, 787–792.
- Felsen, G., Shen, Y.-S., Yao, H., Spor, G., Li, C., & Dan, Y. (2002). Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron*, *36*, 945–954.
- Finlayson, P. G., & Cynader, M. S. (1995). Synaptic depression in visual cortex tissue slices: An in vitro model for cortical neuron adaptation. *Experimental Brain Research*, *106*, 145–155.
- Giaschi, D., Douglas, R., Marlin, S., & Cynader, M. S. (1993). The time course of direction-selective adaptation in simple and complex cells in cat striate cortex. *Journal of Neurophysiology*, *70*, 2024–2034.
- Greenlee, M. W., Georgeson, M. A., Magnussen, S., & Harris, J. P. (1991). The time course of adaptation to spatial contrast. *Vision Research*, *31*, 223–236.
- Greenlee, M. W., & Magnussen, S. (1987). Saturation of the tilt aftereffect. *Vision Research*, *27*, 1041–1043.
- Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*, 14–23.
- Heeger, D. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, *9*, 181–197.
- Heeger, D., Simoncelli, E. P., & Movshon, J. A. (1996). Computational models of cortical visual processing. *Proceedings of the National Academy of Sciences, USA*, *93*, 623–627.
- Hyvärinen, A., Hurri, J., & Hoyer, P. O. (2009).

- Natural image statistics* (Vol. 39). London: Springer.
- Ibbotson, M. R. (2005). *Physiological mechanisms of adaptation in the visual system*. In C. W. G. Clifford & G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and after-effects in high-level vision* (pp. 17–46). Oxford, UK: Oxford University Press.
- Jeyabalaratnam, J., Bharmauria, V., Bachatene, L., Cattan, S., Angers, A., & Molotchnikoff, S. (2013). Adaptation shifts preferred orientation of tuning curve in the mouse visual cortex. *PloS One*, *8*, e64294.
- Kayser, C., Einhäuser, W., & König, P. (2003). Temporal correlations of orientations in natural scenes. *Neurocomputing*, *52–54*, 117–123.
- King, J. L., Lowe, M. P., & Crowder, N. A. (2015). Contrast adaptation is spatial frequency specific in mouse primary visual cortex. *Neuroscience*, *310*, 198–205.
- Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. New York: Cambridge University Press.
- Kohn, A. (2007). Visual adaptation: Physiology, mechanisms, and functional benefits. *Journal of Neurophysiology*, *97*, 3155–3164.
- Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: From single cells to BOLD signals. *Trends in Neuroscience*, *29*, 250–256.
- Kurahashi, T., & Menini, A. (1997). Mechanism of odorant adaptation in the olfactory receptor cell. *Nature*, *385*, 725–729.
- Ledue, E. E., King, J. L., Stover, K. R., & Crowder, N. A. (2013). Spatiotemporal specificity of contrast adaptation in mouse primary visual cortex. *Frontiers in Neural Circuits*, *7*, 154.
- Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences, USA*, *96*, 10530–10535.
- Lochmann, T., Ernst, U. A., & Deneve, S. (2012). Perceptual inference predicts contextual modulations of sensory responses. *Journal of Neuroscience*, *32*, 4179–4195.
- Maffei, L., Fiorentini, A., & Bisti, S. (1973). Neural correlate of perceptual adaptation to gratings. *Science*, *182*, 1036–1038.
- Maravall, M., Petersen, R. S., Fairhall, A. L., Arabzadeh, E., & Diamond, M. E. (2007). Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS Biology*, *5*, e19.
- Movshon, J. A., & Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurones. *Nature*, *278*, 850–852.
- Müller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. *Science*, *285*, 1405–1408.
- Müller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (2001). Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience*, *21*, 6978–6990.
- Nelson, S. B. (1991). Temporal interactions in the cat visual system. III. Pharmacological studies of cortical suppression suggest a presynaptic mechanism. *Journal of Neuroscience*, *11*, 369–380.
- Ohzawa, I., Sclar, G., & Freeman, R. D. (1982). Contrast gain control in the cat visual cortex. *Nature*, *298*, 266–268.
- Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cat's visual system. *Journal of Neurophysiology*, *54*, 651–667.
- Olshausen, B. A., & Field, D. J. (2000). Vision and the coding of natural images. *American Scientist*, *88*, 238–245.
- Parra, L., Spence, C., & Sajda, P. (2001). Higher-order statistical properties arising from the non-stationarity of natural signals. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Nips* (pp. 786–792). Cambridge, MA: MIT Press.
- Patterson, C. A., Duijnhouwer, J., Wissig, S. C., Krekelberg, B., & Kohn, A. (2014). Similar adaptation effects in primary visual cortex and area MT of the macaque monkey under matched stimulus conditions. *Journal of Neurophysiology*, *111*, 1203–1213.
- Patterson, C. A., Wissig, S. C., & Kohn, A. (2013). Distinct effects of brief and prolonged adaptation on orientation tuning in primary visual cortex. *Journal of Neuroscience*, *33*, 532–543.
- Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: An overview. *Frontiers in Integrative Neuroscience*, *8*, 19.
- Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–70.
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87.

- Sanchez-Vives, M. V., Nowak, L. G., & McCormick, D. A. (2000). Cellular mechanisms of long-lasting adaptation in visual cortical neurons in vitro. *Journal of Neuroscience*, *20*, 4286–4299.
- Saul, A. B., & Cynader, M. S. (1989). Adaptation in single units in visual cortex: The tuning of aftereffects in the temporal domain. *Visual Neuroscience*, *2*, 609–620.
- Schwartz, O., Hsu, A., & Dayan, P. (2007). Space and time in visual context. *Nature Reviews. Neuroscience*, *8*, 522–535.
- Schwartz, O., Sejnowski, T. J., & Dayan, P. (2006). Soft mixer assignment in a hierarchical generative model of natural scene statistics. *Neural Computation*, *18*, 2680–2718.
- Schwartz, O., Sejnowski, T. J., & Dayan, P. (2009). Perceptual organization in the tilt illusion. *Journal of Vision*, *9*(4):19, 1–20, doi:10.1167/9.4.19. [PubMed] [Article]
- Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, *4*, 819–825.
- Scialoja, G., Lennie, P., & DePriest, D. D. (1989). Contrast adaptation in striate cortex of macaque. *Vision Research*, *29*, 747–755.
- Seriès, P., Stocker, A. A., & Simoncelli, E. P. (2009). Is the homunculus “aware” of sensory adaptation? *Neural Computation*, *21*, 3271–3304.
- Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*, 936–942.
- Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature*, *378*, 492–496.
- Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multiscale transforms. *IEEE Transactions on Information Theory*, *38*, 587–607.
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216.
- Sinz, F., & Bethge, M. (2013). Temporal adaptation enhances efficient contrast gain control on natural images. *PLoS Computational Biology*, *9*, e1002889.
- Solomon, S. G., & Kohn, A. (2014). Moving sensory adaptation beyond suppressive effects in single neurons. *Current Biology*, *24*, R1012–R1022.
- Spratling, M. W. (2012). Unsupervised learning of generative and discriminative weights encoding elementary image components in a predictive coding model of cortical function. *Neural Computation*, *24*, 60–103.
- Stevenson, I. H., Cronin, B., Sur, M., & Kording, K. P. (2010). Sensory adaptation and short term plasticity as Bayesian correction for a changing brain. *PloS One*, *5*, e12436.
- Stocker, A. A., & Simoncelli, E. P. (2006). Sensory adaptation within a Bayesian framework for perception. In Y. Weiss, B. Schölkopf, & J. C. Platt (Eds.), *Nips*, (pp. 1289–1296). Cambridge, MA: MIT Press.
- Teich, A. F., & Qian, N. (2003). Learning and adaptation in a recurrent model of V1 orientation selectivity. *Journal of Neurophysiology*, *89*, 2086–2100.
- Trott, A. R., & Born, R. T. (2015). Input-gain control produces feature-specific surround suppression. *Journal of Neuroscience*, *35*, 4973–4982.
- Turner, R. E., & Sahani, M. (2008). Modeling natural sounds with modulation cascade processes. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), *Nips* (pp. 1545–1552). Cambridge, MA: MIT Press.
- Ursino, M., Magosso, E., & Cuppini, C. (2008). Possible mechanisms underlying tilt aftereffect in the primary visual cortex: A critical analysis with the aid of simple computational models. *Vision Research*, *48*, 1456–1470.
- Vautin, R. G., & Berkley, M. A. (1977). Responses of single cells in cat visual cortex to prolonged stimulus movement: neural correlates of visual aftereffects. *Journal of Neurophysiology*, *40*, 1051–1065.
- Vul, E., Krizay, E., & MacLeod, D. I. A. (2008). The McCollough effect reflects permanent and transient adaptation in early visual cortex. *Journal of Vision*, *8*(12):4, 1–12, doi:10.1167/8.12.4. [PubMed] [Article]
- Wainwright, M. J., Schwartz, O., & Simoncelli, E. P. (2002). Natural image statistics and divisive normalization: Modeling nonlinearity and adaptation in cortical neurons. In R. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 203–222). Cambridge, MA: MIT Press.
- Wainwright, M. J., & Simoncelli, E. P. (2000). Scale mixtures of Gaussians and the statistics of natural images. In S. A. Solla, T. K. Leen, and K. R. Müller (Eds.), *Nips* (pp. 855–861). Cambridge, MA: MIT Press.

- Wark, B., Fairhall, A. L., & Rieke, F. (2009). Timescales of inference in visual adaptation. *Neuron*, *61*, 750–761.
- Webb, B. S., Dhruv, N. T., Solomon, S. G., Tailby, C., & Lennie, P. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *Journal of Neuroscience*, *25*, 11666–11675.
- Webster, M. A. (2011). Adaptation and visual coding. *Journal of Vision*, *11*(5):3, 1–23, doi:10.1167/11.5.3. [PubMed] [Article]
- Weiss, Y., & Adelson, E. H. (1998). Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. *Center for Biological and Computational Learning Paper*, *158*, 1–42.
- Wilson, D. A. (2009). Olfaction as a model system for the neurobiology of mammalian short-term habituation. *Neurobiology of Learning and Memory*, *92*, 199–205.
- Wissig, S. C., & Kohn, A. (2012). The influence of surround suppression on adaptation effects in primary visual cortex. *Journal of Neurophysiology*, *107*, 3370–3384.
- Wolfe, J. M., & O’Connell, K. M. (1986). Fatigue and structural change: Two consequences of visual pattern adaptation. *Investigative Ophthalmology & Visual Science*, *27*, 538–543. [PubMed] [Article]

Appendix A

Here we provide further implementation details for the binary MGSM. The excitatory drive to the RF of a neuron with preferred orientation θ is given by $x_{t,\theta}$. The normalization pool includes the four possible RF orientations in the present, namely, $x_t = (x_{t,0}, x_{t,45}, x_{t,90}, x_{t,135})$, and only the preferred orientation θ in the past, namely, $x_p = (x_{p_1,\theta}, x_{p_2,\theta}, \dots, x_{p_8,\theta})$, where the indexes p_1, p_2, \dots, p_8 denote the eight past frames. This binary model implements neuron-specific normalization, as the tuning of the normalization signal is determined exclusively by the tuning of the neuron’s RF.

In the binary MGSM, the estimate of the Gaussian component in the present time is

$$E[g_{t,\theta}|x_t, x_p] = p(\xi_1|x_t, x_p)E[g_{t,\theta}|x_t, \xi_1] + p(\xi_2|x_t, x_p)E[g_{t,\theta}|x_t, x_p, \xi_2] \quad (5)$$

In Equation 5, we have introduced a binary variable with values ξ_1 and ξ_2 , for the independent and dependent cases, respectively. The first factor of each term on the right-hand side of Equation 5 is the

(posterior) probability of the input movie being independent or dependent. These factors weight the Gaussian estimate for each of the cases and are obtained by applying Bayes’s rule, that is, combining the prior probability that any given movie sequence is dependent or independent (denoted $\xi_{1\text{prior}}$ and $\xi_{2\text{prior}}$, respectively) with the likelihood that the observed movie was generated by an independent or dependent GSM:

$$p(\xi_1|x_t, x_p) = \frac{\xi_{1\text{prior}}P(x_t, x_p|\xi_1)}{\xi_{1\text{prior}}P(x_t, x_p|\xi_1) + \xi_{2\text{prior}}P(x_t, x_p|\xi_2)}$$

$$p(\xi_2|x_t, x_p) = 1 - p(\xi_1|x_t, x_p) \quad (6)$$

For the second term on the right-hand side of Equation 5, the estimate of the Gaussian component corresponds to the model in which the movie frames are dependent, ξ_2 ; because the present and past are connected by the shared mixer, v , the Gaussian estimate is dependent on the present and the past RFs outputs, x_t, x_p , as in Equation 2. Conversely, in the first term on the right-hand side of Equation 5, the estimate corresponds to the case in which the movie frames are independent, ξ_1 ; because there is no connection between the present and the past, the Gaussian estimate of the present is dependent only on the RF outputs in the present, x_t (see Coen-Cagli et al., 2009; Schwartz et al., 2006, for derivations):

$$E[g_{t,\theta}|x_t] \approx \frac{|x_{t,\theta}|}{\sqrt{\lambda_t}} \quad (7)$$

This is similar to the dependent case (Equation 2); however, the denominator includes only the present frame:

$$\lambda_t = \sqrt{(x_t)^T(\Sigma_t)^{-1}(x_t)} \quad (8)$$

Each GSM component of the MGSM in Equation 5 is a self-contained form of divisive normalization. The λ terms in Equation 2 and Equation 7 normalize the RF output in the present, $x_{t,\theta}$, for the dependent and independent cases, respectively.

Appendix B

Here we provide further details for the multi-orientation MGSM. Given a movie sequence, the past and present RF outputs are assumed to be generated by either an independent GSM (Equation 4) or one of four possible dependent GSMs. The first dependent GSM assumes that a common mixer is shared between the present RFs and the past RF group with orientation 0° , whereas all other past groups have independent mixers:

$$\begin{aligned} x_t &= v g_t \\ x_{p,0} &= v g_{p,0} \\ x_{p,\phi} &= v_\phi g_{p,\phi} \quad \text{for } \phi \neq 0 \end{aligned} \quad (9)$$

Similarly, the other three dependent GSMs assume the common mixer is shared between the present and the past group with orientation 45° , 90° , or 135° . These different components of the multiorientation MGSM are labeled by a variable, with value ξ_1 for the independent GSM and value $\xi_{2,\phi}$ for the GSM, where the dependent past group has orientation ϕ . Therefore, in the multiorientation MGSM, the estimate of the Gaussian component in the present time for the neuron with preferred orientation θ is

$$\begin{aligned} E[g_{t,\theta}|x_t, x_{p,0}, \dots, x_{p,135}] &= p(\xi_1|x_t, x_{p,0}, \dots, x_{p,135})E[g_{t,\theta}|x_t, \xi_1] \\ &\quad + \sum_{\phi} p(\xi_{2,\phi}|x_t, x_{p,0}, \dots, x_{p,135})E[g_{t,\theta}|x_t, x_{p,\phi}, \xi_{2,\phi}] \end{aligned} \quad (10)$$

where $p(\xi_1|x_t, x_{p,0}, \dots, x_{p,135}) + \sum_{\phi} p(\xi_{2,\phi}|x_t, x_{p,0}, \dots, x_{p,135}) = 1$. On the right-hand side of Equation 10, the Gaussian estimate in the first term is identical to Equation 7; the remaining terms have the same form as Equation 2, except each involves a different group of past RFs and their covariance matrix:

$$E[g_t|x_{t_p}^{\theta_i}, \xi_2^{\theta_i}] \approx \frac{|x_t^{\theta_i}|}{\sum_{i=1}^4 \lambda_{t_p}^{\theta_i}} \quad (11)$$

where

$$\lambda_{t_p,\phi} = \sqrt{(x_{p,\phi}, x_t)^T (\Sigma_{t_p,\phi})^{-1} (x_{p,\phi}, x_t)} \quad (12)$$

Similar to the binary MGSM, the probabilities of dependency are estimated using Bayes's rule:

$$\begin{aligned} p(\xi_1|x_t, x_{p,0}, \dots, x_{p,135}) &= \frac{\xi_{1\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_1)}{\xi_{1\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_1) + \sum_{\phi} \xi_{2,\phi\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_{2,\phi})} \\ p(\xi_{2,\phi}|x_t, x_{p,0}, \dots, x_{p,135}) &= \frac{\xi_{2,\phi\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_{2,\phi})}{\xi_{1\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_1) + \sum_{\phi} \xi_{2,\phi\text{prior}} p(x_t, x_{p,0}, \dots, x_{p,135}|\xi_{2,\phi})} \end{aligned} \quad (13)$$