Adaptability and efficiency in neural coding

Wiktor Młynarski[1]* and Ann M. Hermundstad[2]*

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria

[2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

*Equal contribution; for correspondence: wiktor.mlynarski@ist.ac.at, hermundstada@janelia.hhmi.org

Abstract

Animals exhibit remarkable behavioral flexibility, robustly performing demanding tasks —such as searching for food or avoiding predators— in a variety of different contextual and environmental conditions. However, the demands that detecting and adjusting to changes in the environment place on a sensory system often differ from the demands associated with performing a specific behavioral task, even when both objectives rely on the same sensory modality. This necessitates neural encoding strategies that can dynamically balance these conflicting needs. Here, we develop a theoretical framework that explains how this balance can be achieved, and we use this framework to study tradeoffs in speed, performance, and information transmission that arise as a consequence of efficient coding in dynamic environments. This work generalizes current theories of efficient neural coding to dynamic environments, and thereby provides a unifying perspective on adaptive neural dynamics across different sensory systems, environments, and tasks.

INTRODUCTION

Animals must keep up with changes in their sensory environment in order to thrive in diverse surroundings. The barn owl, for example, is able to localize its prey despite changing acoustic properties of the environment [1, 2]; the dragonfly can predict the location of moving insects against different visual backdrops [3]; dogs can track odor plumes over large distances that span multiple odor environments [4]. In order to accurately perform such behavioral tasks, animals must dynamically adapt to changes in their environment. But detecting and adapting to environmental changes often relies on different stimulus features than searching for food or avoiding predators. Given limited resources for supporting such tasks, the brain must not only prioritize some stimulus features over others, it must also dynamically change this prioritization as the environment and the behavioral demands change.

These changing priorities are thought to be reflected in the changing response properties of neurons across the brain. In early sensory areas, for example, neurons are known to adapt their dynamic range to changes in low-order features of the environment, such as the mean [5] or contrast [6–8] of incoming stimuli. In deeper brain areas like the hippocampus and entorhinal cortex, cells modify their firing fields based on changes in high-order features, such as the locations of rewards in the environment [9, 10]. Numerous other studies have similarly characterized changes in the response properties of individual neurons [6, 11] and neural populations [7, 12, 13] to changes in simple [5–8] and complex [12, 14] features of the environment, and across visual [15], auditory [7, 16, 17], olfactory [18–20], somatosensory [21], and electrosensory [22] domains.

A dominant hypothesis is that this diversity of adaptive phenomena reflects a common underlying principle, as posited within the framework of efficient coding [23]: First, sensory systems have limited metabolic resources, and should therefore make efficient use of these resources by prioritizing those stimulus features that are relevant for the task at hand [24–26]. Second, when the environment changes, sensory systems should adapt in order to prioritize those features that are relevant under the new environmental conditions [5, 6, 11, 27]. These two statements, however, cannot be consistently resolved within the same framework; by prioritizing features that are relevant for a given task, sensory systems can be blind to features that signal changes in the environment and that would prompt the system to flexibly adapt to these changes. Indeed, while efficient coding can account for differences in coding properties before and after a change in stimulus statistics [5, 6, 11, 27], it cannot explain how they come about.

In order to flexibly adapt to environmental changes, sensory systems should dynamically balance task performance with the ability to detect changes in the environment. As we will show, failure to balance these two objectives can lead to prolonged periods of poor performance that could manifest in neural dynamics, and ultimately behavior. We develop a new theoretical framework for understanding how this balance could be achieved, and we use

this framework to study tradeoffs in speed and performance across a range of different tasks and sensory environments. To our knowledge, this work is the first to provide a normative account of neural coding dynamics underlying adaptive phenomena in the brain.

RESULTS

A theoretical framework for adaptation.
Animals must remain flexible to changes in the environment in order to meet their behavioral demands. As a simple example, consider a system whose goal is to accurately detect the presence of a predator. Different environments might be occupied by different predators, which might in turn be signaled by different distributions of stimulus features (Fig. 1A). A successful system should have a neural code that can discriminate stimuli within these different distributions. A simple neural code could be constructed by transforming incoming stimuli through a saturating nonlinearity in a manner that approximates the stochastic response function of a sensory neuron (Fig. 1B). Intuitively, this neuron should use its limited dynamic range to discriminate those stimuli that are most likely to occur in its current environment; this can be achieved by aligning the steep part of its nonlinear response function with the bulk of the incoming stimulus distribution (Fig. 1B, left). This alignment guarantees that the neuron's limited coding capacity is efficiently allocated for the task of accurate stimulus reconstruction, provided that the incoming stimulus distribution does not change in time [26, 28, 29].

In principle, if the environment were to change, the neuron should shift its response function to align with the new distribution in order to maintain accurate performance. In practice, however, the neuron might be insensitive to the stimuli that signal the change, and might then fail to adapt and consequently be unable to distinguish stimuli that signal a predator in the new environment (Fig. 1B, right). Whether and how the neuron adapts to this change, and the impact that this adaptation has on the speed and accuracy of its performance, depends on how the system prioritizes task performance (in this case, reconstructing stimuli to detect a predator) relative to detecting changes in the distribution of stimuli that underlie this task. The balance between these objectives can lead to different types of failures, including catastrophic errors and prolonged periods of error (Fig. 1C).

To explore the dynamic interplay between these objectives, we consider a system with an adaptable neural encoding (the nonlinear response function in Fig. 1B is one such example encoding) whose output is used to perform different computational tasks on incoming sensory stimuli (Fig. 1D). The encoding is constrained by the precision with which incoming stimuli are represented in neural responses. It must therefore devote its limited coding capacity to accurately representing task-relevant stimuli while simultaneously enabling the system to detect changes in the underlying distribution of these stimuli. These two objectives are used to dynamically adapt the encoding based on an internal estimate of the stimulus distribution. The balance between these two objectives will be manifested in the dynamics of both the neural encoding and task errors (Fig. 1E).

To understand the relationship between these dynamics, we adopt the framework of rate-distortion (RD) theory [25, 30, 31]. This framework relies on two fundamental quantities: the mutual information (or 'rate') between the stimulus and the neural response, and the error (or 'distortion') in task performance induced by inaccuracies of the sensory encoding. The mutual information quantifies how many features of the stimulus (measured in bits) are encoded in the neural response. The error in task performance reflects the fact that only certain task-relevant features of the stimulus are encoded. RD theory rigorously specifies an optimal relationship between these two quantities in the form of an RD curve (Fig. 1F, solid black line). Every point on the RD curve determines the minimum task error that can be achieved by a system whose capacity limits the maximum number of stimulus features that it can encode. When the environment is in a steady-state, a sensory system that uses its full capacity to encode only task-relevant stimulus features is said to be optimally adapted to the distribution of incoming stimuli, and as such is described by a single point on the RD curve (Fig. 1F, black marker).

If the environment changes abruptly, a code that was previously adapted will no longer convey relevant features of the new stimulus distribution. This results in an abrupt increase in task error and is marked by a movement away from the RD curve (Fig. 1F, dashed arrows). In this work, we characterize adaptation as the process of returning to the point on the RD curve that represents an optimal solution in the new environment (Fig. 1F, solid arrows). During this process, when the system does not lie on the RD curve, the system is not adapted and is therefore making suboptimal use of its limited capacity. There are two possible sources of this suboptimality (Fig. 1F, bottom right). First, the system can be under-using its capacity by encoding fewer stimulus features than its capacity would allow. Second, the system can be misusing its capacity by encoding features of the stimulus that are not relevant for the

task at hand. Both scenarios lead to suboptimal task performance due to a reduction in task-relevant information. Crucially, during adaptation, task error could exceed levels critical for survival (as in the example of detecting a predator), and prolonged periods of adaptation could expose the system to this risk for an extended period of time. We argue that the speed of adaptation and the maximum task error are fundamental quantities for characterizing the dynamics of adaptation from an information-processing perspective (Fig. 1F, top right).

In what follows, we examine the dynamics of adaptation for different encoding schemes (Fig. 1D-II) that must support different task-relevant computations (Fig. 1D-III) in environments with different stimulus statistics (Fig. 1D-I). In all cases, the system must maintain and update an internal estimate of the stimulus distribution (Fig. 1D-IV) in order to dynamically adapt the encoding. We compare adaptation strategies in which the encoding is optimized for either task performance (i.e., minimizing task error) or adaptability (i.e., maximizing the detection of changes in the stimulus distribution), and we construct interpolated codes that can leverage the advantages of each strategy.

The dynamic interplay between performance and adaptability.

Adaptation is commonly described by changes in the tuning properties of sensory neurons in response to changes in low-order stimulus statistics (e.g. [8, 32–34]). We use this scenario to build intuition about the general principles that govern adaptive neural dynamics. To this end, we model a neuron that is designed to reconstruct stimuli drawn from a Gaussian stimulus distribution whose mean (Fig. 2-I) or standard deviation (Fig. 3-I) can change abruptly over time between two values. The stochastic spiking response of this neuron is determined by a parameterized nonlinearity (Figs. 2A-II, 3A-II) and is used to linearly decode the stimulus (Figs. 2A-III, 3A-III). A downstream observer then uses this stimulus estimate to update an internal estimate of the mean or standard deviation of the stimulus distribution (Figs. 2A-IV, 3A-IV), which is then fed back upstream and used to adapt the encoding nonlinearity.

The parameters of the encoding nonlinearity determine how accurately the system can reconstruct incoming stimuli, and how quickly it can detect changes in the distribution from which these stimuli are drawn. When the mean of the stimulus distribution is changing in time (Fig. 2), a nonlinearity optimized for reconstruction maintains a fixed slope, and its offset aligns with the internal estimate of the stimulus mean (purple curves in Fig. 2B,D; Methods). Just before a change, this internal estimate is accurate, and the offset of the nonlinearity is closely aligned with true mean of the stimulus distribution (Fig. 2B). Just after a change, the bulk of incoming stimuli fall within the saturating regime of the nonlinearity. This results in a saturated firing rate (purple curve in Fig. 2H) that prevents the system from accurately reconstructing those stimuli that signal the change (purple curve in Fig. 2G). This, in turn, significantly slows adaptation (purple curve in Fig. 2F). Note that the firing rate dynamics are asymmetric to increases versus decreases in the stimulus mean; a simple parameter-free recoding procedure, known as entropy coding [35], generates dynamics that require fewer spikes to maintain the same levels of accuracy and that more closely resemble those observed experimentally (insets of Fig. 3H; Methods).

A nonlinearity that is instead optimized for adaptability should enable rapid detection of changes in the mean of the stimulus distribution. Such a nonlinearity maintains a broad slope, and its offset is shifted away from the mean of the incoming stimulus distribution (Fig. 2C-D). When the mean changes abruptly, incoming stimuli do not fully saturate the output of the nonlinearity, and can therefore be accurately reconstructed (green curve in Fig. 2G). These stimuli induce uncertainty in the observer's estimate of the stimulus mean. To resolve this uncertainty, the nonlinearity transiently sharpens and shifts to the midpoint between the two stimulus distributions (Fig. 2C). This enables quick discrimination of stimuli that differentiate the two distributions, which in turns leads to rapid adaptation to the new mean (green curves in Fig. 2F-G). This transient period of adaptation is signaled by a small and brief change in firing rate (green curve in Fig. 2H).

We observe qualitatively similar behavior when the variance of the stimulus distribution is changing in time (Fig. 3). Here, a nonlinearity optimized for stimulus reconstruction has an offset that remains fixed over time and a slope that tracks the estimated standard deviation of the incoming stimulus distribution (purple curves in Fig. 3B,D; Methods). This enables the system to accurately resolve those stimuli that are likely given the system's current estimate of the stimulus distribution. In contrast, a nonlinearity optimized for adaptability retains resolution for those stimuli that would signal a change (green curves in Fig. 3C-D). Because such stimuli can be either likely or unlikely under the current estimate of the stimulus distribution, we see asymmetries in response dynamics following an increase or decrease in stimulus variance (Fig. 3F-G). Despite their differences, both strategies produce constant firing rates (Fig. 3H), a finding that would be predicted for any symmetric nonlinearity that is centered on the (symmetric) stimulus distribution. As in the case of mean adaptation, entropy coding leads to an overall reduction in firing rates and produces transient dynamics that again more closely resemble experimental observations (insets

of Fig. 3H; compare to [36]).

On individual trials, nonlinearities exhibit the same qualitative changes in slope and offset, but these changes are larger in magnitude and shorter in duration than is observed on average (Fig. 2E, 3E). Because these changes occur at varying temporal offsets from a switch in the stimulus distribution, abrupt changes in single-trial dynamics can lead to the smooth and gradual adaptation observed on average.

Tradeoffs between steady-state and transient performance.

To understand dynamics of information processing that underlie these observable changes in firing rates and tuning curves, we use rate-distortion (RD) theory to quantify the mutual information between the input and output of the neural encoding (i.e., the mutual information between incoming stimuli and spike counts), and the error that this encoding induces in task performance (i.e., the error in stimulus reconstruction). The dynamically changing relationship between these two quantities generates a trajectory on the RD plane (Fig. 4A,C,D).

Projecting dynamics on the RD plane enables a principled definition of adaptation as the return to a stationary point on the RD plane, which coincides with the time that the system reaches a stable internal estimate of the stimulus distribution. In steady state, when the stimulus distribution is stationary, an optimal adapted system is represented as a single point on the plane. When comparing a system optimized for stimulus reconstruction versus adaptability, we see that the former maintains lower reconstruction error in steady-state (insets of Fig. 4A,C,D). When the environment abruptly changes, both systems are "kicked" from their steady-state locations (dotted arrows) and must dynamically adapt in order to return (solid lines). The dynamics of this return depend on whether there was a change in the mean (Fig. 4A) or variance (Fig. 4C-D) of the stimulus distribution, and in the latter case, whether this change led to an increase (Fig. 4C) or decrease (Fig. 4D) in variance.

The details of trajectories on the RD plane further enable a principled analysis of changes in information processing that occur during the transient period of adaptation. Following a change in the mean of the stimulus distribution, a nonlinearity optimized for reconstruction transmits very little information about incoming stimuli. This drop in information leads to high errors in task performance and significantly slows the system in adapting to this change (purple curve in Fig. 4A). A nonlinearity optimized for adaptability can improve transient performance by compromising information transmission, and thus performance, in steady state (green curves in Fig. 4A).

Together, this analysis reveals tradeoffs between steady-state and transient performance (Fig. 4B) that could not be gleaned from firing rates or tuning curves alone. A system optimized for reconstruction maintains lower error in steady state, but suffers from long adaptation times and high transient errors when the stimulus distribution changes. In contrast, a system optimized for adaptability is able to quickly and accurately respond to changes in the stimulus distribution, but suffers from higher errors when the stimulus distribution is stationary. We can leverage the advantages of each strategy by designing interpolated systems that balance task performance and adaptability. These systems generate trajectories that occupy an intermediate region of the RD plane (turquoise curves in Fig. 4A, C, D), and that are accompanied by adaptive changes in tuning curves and firing rates (turquoise curves in Figs. 2D,H and 3D,H). The degree of these tradeoffs depends on environmental statistics; across all strategies, we find that the balance between speed and error is impacted by both the variability and volatility of the environment (defined by the magnitude of change and the rate of change in the stimulus distribution, respectively; open markers in Fig. 4B).

Adaptive population coding of complex stimulus features.

These same principles can be generalized to neural populations, where adaptive phenomena have been observed experimentally [12, 13] and can be manifested by a shifting of neural tuning curves in response to changes in stimulus statistics [37]. To illustrate adaptation in population-level dynamics, we model a population of noisy neurons that is designed to accurately reconstruct stimuli drawn from two different distributions of spatial frequency derived from video footage of animals in the African savanna [38] (Fig. 5AI-AIII). Each neuron in the population is characterized by a Gaussian tuning curve that specifies the neuron's average firing rate in response to incoming stimuli. We assume tuning curves with identical widths and identical maximum firing rates, but with variable preferred locations along the stimulus axis. The set of locations is determined by a linear density function that maps uniformly-spaced values on the interval $[0, 1]$ to locations along the stimulus axis [26, 39] (Fig. 5A-II). Manipulating the slope and offset of this density function produces different distributions of tuning curves that determine the precision with which incoming stimuli are encoded in the population response. This parametrization provides a simple and compact description of the entire population, regardless of its size.

In order to support optimal reconstruction, the population of tuning curves should uniformly tile the range spanned by the incoming stimulus distribution (Fig. 5B, top row). When the distribution changes, the stimulus estimates decoded from the population response show high errors (purple curve in Fig. 5C) that persist until a downstream observer can correctly infer the change (purple curve in Fig. 5D) and can redistribute the tuning curves (Methods). In order to facilitate rapid detection of this change, a maximally adaptive system should distribute its tuning curves in a manner that optimally discriminates stimuli from the two distributions. Because the stimulus distributions do not have a parametric form, this is equivalent to a classification task that can be implemented by maximizing the symmetrized Kullback-Leibler divergence between the distributions of population responses obtained for each stimulus distribution (Methods). The resulting tuning curves are concentrated close to the peak of each stimulus distribution (Fig. 5B, bottom row), and enable the downstream observer to quickly infer a change from the population response (green curve in Fig. 5D). This, in turn, maximizes the speed of adaptation and minimizes the impact on reconstruction error during the transient period of adaptation (green curves in Fig. 5CD). By constructing populations that are optimized for a combination of reconstruction and adaptability (Fig. 5B, middle row), it is possible to generate a family of intermediate codes that navigate a tradeoff between performance in steady state, and the speed and accuracy of performance during the transient period of adaptation (Fig. 5E).

Adaptive predictive coding in the face of noise.
While many studies have characterized adaptive changes in neural tuning curves, time-varying stimulus statistics can also trigger adaptive changes in more complex properties of neural responses, such as the structure of receptive fields [12, 40] and predictive filters [41]. To explore how such structure should optimally adapt over time, we model a neuron whose response is characterized by a center-surround receptive field that is designed to predict the central pixel of an image patch from the surrounding pixels [42] (Fig. 6A-II). The neuron conveys downstream only the residual error between this prediction and the value of the central pixel (Fig. 6A-III), which reduces the energetic cost of stimulus coding [42].

The shape of the optimal predictive filter depends on the level of noise in the image patches [42] (Fig. 6B, top row). We consider a scenario in which this noise can fluctuate between a low and high level over time (Fig. 6A). When the noise level changes, the neuron must dynamically switch its predictive filter to maintain accurate predictions under the new noise conditions. However, a downstream observer is very slow to infer this change from the output of a filter that is optimized for the wrong noise conditions (purple curve in Fig. 6D). This leads to transiently high prediction errors when the noise level changes (purple curve in Fig. 6C), which persist until the system can correctly infer this change and update its filter accordingly. Alternatively, by using a filter that is optimized for adaptability, the system can quickly infer changes in the noise level, but suffers from higher prediction errors when the environment is stable (green curves in Fig. 6C-D). By designing filters that interpolate between optimal prediction and adaptability (Fig. 6B, third row from the top), the system can increase its speed of adaptation at the cost of reduced task performance in steady state (Fig. 6E). Due to the asymmetric nature of the task, the same interpolated filters can cause increases in maximum prediction error when the noise level increases, but decreases in error when the noise level decreases.

DISCUSSION

Adaptive phenomena in neural coding have long been a subject of experimental and theoretical research. To date, the dynamics of adaptation have been traditionally characterized in terms of changes in experimentally-observable quantities such as firing rates, tuning curves, and receptive fields [7, 8, 22, 27, 43–46], and a majority of modeling work has focused on reproducing such dynamics in specific sensory systems [36, 47–50]. While such models can provide useful summaries of experimental phenomena, they do not explain the role that these dynamics might serve in performing task-relevant computations, nor do they explain whether these diverse phenomena share common design principles. Such questions fall within the domain of normative approaches to neural coding. Existing normative accounts of adaptation, largely developed within the framework of efficient coding, propose that adaptation readjusts the sensory code to maintain a constant rate of information transmission [6, 11, 12, 27, 51]. Prominent examples are found in the retina, which has been shown to adapt its code to maintain efficient information transmission after a change in stimulus statistics [41] and in a manner that depends on cell type [40]. Adaptation in the fly visual system has also been analyzed in terms of information transmission [52], and the timescale of adaptation has been compared to the timescale required to infer changes in the stimulus distribution [6, 32]. None of these studies, however,

explicitly treat performance or information processing during the period of adaptation itself, when the sensory code is hypothesized to be dynamically readjusting.

One obstacle to leveraging these different theoretical approaches has been the lack of a conceptual framework for characterizing neural coding during periods of adaptation. We propose that rate-distortion (RD) theory could address this gap by providing a common reference frame for analyzing and comparing the dynamics of adaptive phenomena as trajectories on the RD plane. Here, we go beyond the static formulation of RD theory, where the steady-state performance of a system is described by a single point on the RD plane, to define adaptation as the time during which a system moves between points on the RD plane in response to changes in the environment. To maintain high levels of task performance in the face of these changes, transitions on the RD plane should be rapid and should have minimal impact on performance. The maximum task error and the length of the RD trajectory provide a natural description of task performance and coding efficiency during such periods of adaptation.

To highlight the generality of our approach, we optimized and analyzed several qualitatively different types of neural codes that were designed to perform different computational tasks in changing environments. We found a consistent tradeoff between the accuracy of task performance when the environment was stable, and the speed of adaptation when the environment was changing. This tradeoff arises because the stimuli that support accurate task performance often differ from those that signal changes in the underlying stimulus distribution. Intuitively, a system designed to minimize task error should allocate its limited coding capacity to accurately representing stimuli that occur most frequently. Changes in the distribution of these stimuli, however, are typically signaled by infrequent stimuli that become indistinguishable to the same encoder. As a consequence, a system optimized solely for task performance loses the ability to quickly detect and respond to these changes, which often leads to poor task performance when the environment changes. In order to increase adaptability, the encoder must devote some of its limited capacity to accurately representing stimuli that signal informative environmental changes, and in doing so must deviate from the optimal steady-state performance specified by RD theory. Such deviations from optimality have been observed in multiple systems, including the fly visual system [26, 28] and the retina [25], and could arise as a consequence of balancing task performance with adaptability. Achieving such a balance could rely on task-dependent feedback from higher brain areas [53, 54], or could be implemented more locally via biochemical interactions [55] or dendritic nonlinearities [56]. Regardless of the implementation, the ability to convey task-relevant information in the face of environmental change will depend on which stimuli are represented with high accuracy when these environmental changes occur.

Task-relevant information processing has also been studied within the framework of Information Bottleneck (IB) [24, 25, 57], which is related to RD theory but with the distortion measure replaced by the mutual information between the neural response and a task-relevant variable. Temporal trajectories on the IB plane were recently proposed as a framework for studying learning in artificial neural networks [58]. The similarity to our approach raises the intriguing possibility that adaptation and learning, while occurring on different timescales, might be studied with similar conceptual tools. When viewed through the RD framework, the observed relationship between steady-state performance, transient performance, and the speed of adaptation is reminiscent of fundamental physical tradeoffs between speed, energy, and accuracy that have been described in sensory systems [59, 60]. Taken together, our results fit within a broader spectrum of work that extends existing normative theories of neural coding by synthesizing different—and often competing—objectives such as efficiency, inference [39, 54, 61, 62], and prediction [57, 63].

Since its formulation, the efficient coding hypothesis has been a guiding principle in the study of neural coding [23]. It has generated multiple, now canonical, normative accounts of neural coding by deriving optimal solutions for conveying information about stationary stimulus distributions (e.g. [28, 42, 64, 65]). Natural environments, however, are constantly changing, which is reflected in the dynamic and adaptive structure of sensory codes in the brain. In this work, we developed a normative framework based on rate-distortion theory and dynamic Bayesian inference to study adaptive neural computation. To our knowledge, this is the first systematic extension of the efficient coding hypothesis to nonstationary environments (however, see [66]), and as such enables a normative description of a vast range of sensory adaptation dynamics that were previously outside of the domain of the efficient coding framework.

Sensory adaptation takes many forms and is ubiquitous across the nervous system [7, 8, 12, 16, 22, 32, 40]. The apparent diversity of adaptive phenomena might create the perception that it is not possible to perform meaningful comparisons across different sensory systems or modalities. By proposing a theoretical description that captures the dynamics of information processing during adaptation, our approach facilitates the comparison of these diverse phenomena using a common set of principles. Rigorously connecting this framework to experimental data will require the development of statistical models that can capture time-varying changes in neural nonlinearities and receptive fields [34, 67, 68], and theoretical approaches to deepen our understanding of rate-distortion theory in

non-stationary environments [69]. Together, these developments may shed light on how adaptability and efficiency guide the evolutionary design of neural representations, and thereby provide a normative understanding of a broad range of dynamic phenomena in adaptive neural coding.

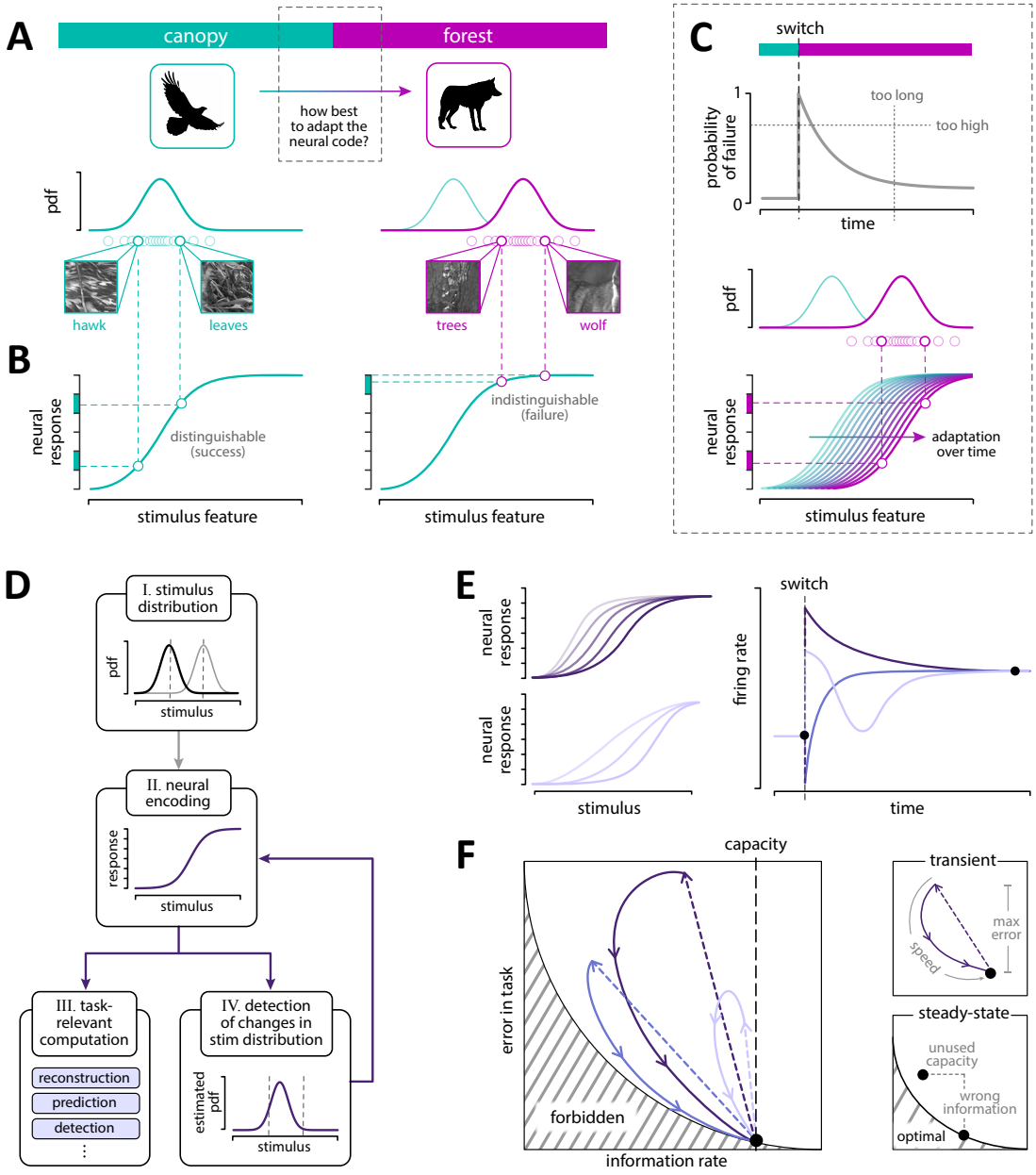

Figure 1: A framework for characterizing the dynamics of adaptation in sensory neurons. A. (Schematic) Animals perform different tasks in changing environments. For example, to detect a predator, an animal might have to discriminate between stimuli (here, local regions of a visual scene) that signal a hawk in the canopy, or a wolf in the forest. **B.** (Schematic) A model neuron encodes these stimuli via a saturating nonlinear response function that generates a discrete number of discriminable responses. In order to discriminate stimuli that signal the presence of a predator, the response function of the neuron should be aligned with the stimulus distribution. When the stimulus distribution changes (right column), incoming stimuli might no longer be discriminable to the same neuron, and the neuron could fail to signal the presence of a predator. **C.** (Schematic) In order to accurately signal a predator in the new environment, the neuron should adapt its response function (lower panel). The dynamics of this adaptation will determine the speed and accuracy of performance, and the degree and duration of errors (top panel). **D.** (Schematic) We formulate a general framework for an adaptive neural system. Stimuli originating from a time-varying stimulus distribution (I) are encoded via a neural response function (II). The output of this encoding is used both to solve diverse computational tasks (III), and to track changes in the stimulus distribution that underlies these tasks (IV). An estimate of the stimulus distribution is used to dynamically adapt the encoding over time. **E.** (Schematic) The dynamics of adaptation will determine the detailed changes in the response function of the encoder (left), and the firing rate dynamics (right). **F.** (Schematic) The dynamics of adaptation can be characterized as a trajectory on the rate-distortion (RD) plane. The RD plane relates the information that encoding conveys about incoming stimuli, and the error that this encoding induces in the task of stimulus reconstruction. An optimally-adapted system is marked by a single point (black marker) on the RD curve (black line). A change in the stimulus distribution will "kick" the system away from the RD curve (dashed arrows), and the system must dynamically adapt its response function in order to return (solid arrows). Different response function dynamics (panel E) give rise to different trajectories on the RD plane. Lower right: Departures from an optimal point on the RD curve can reflect both underuse and misuse of coding capacity. Upper right: Individual trajectories can be characterized by their curvilinear length (quantifying the speed of adaptation), and their maximum height (quantifying the maximum error).

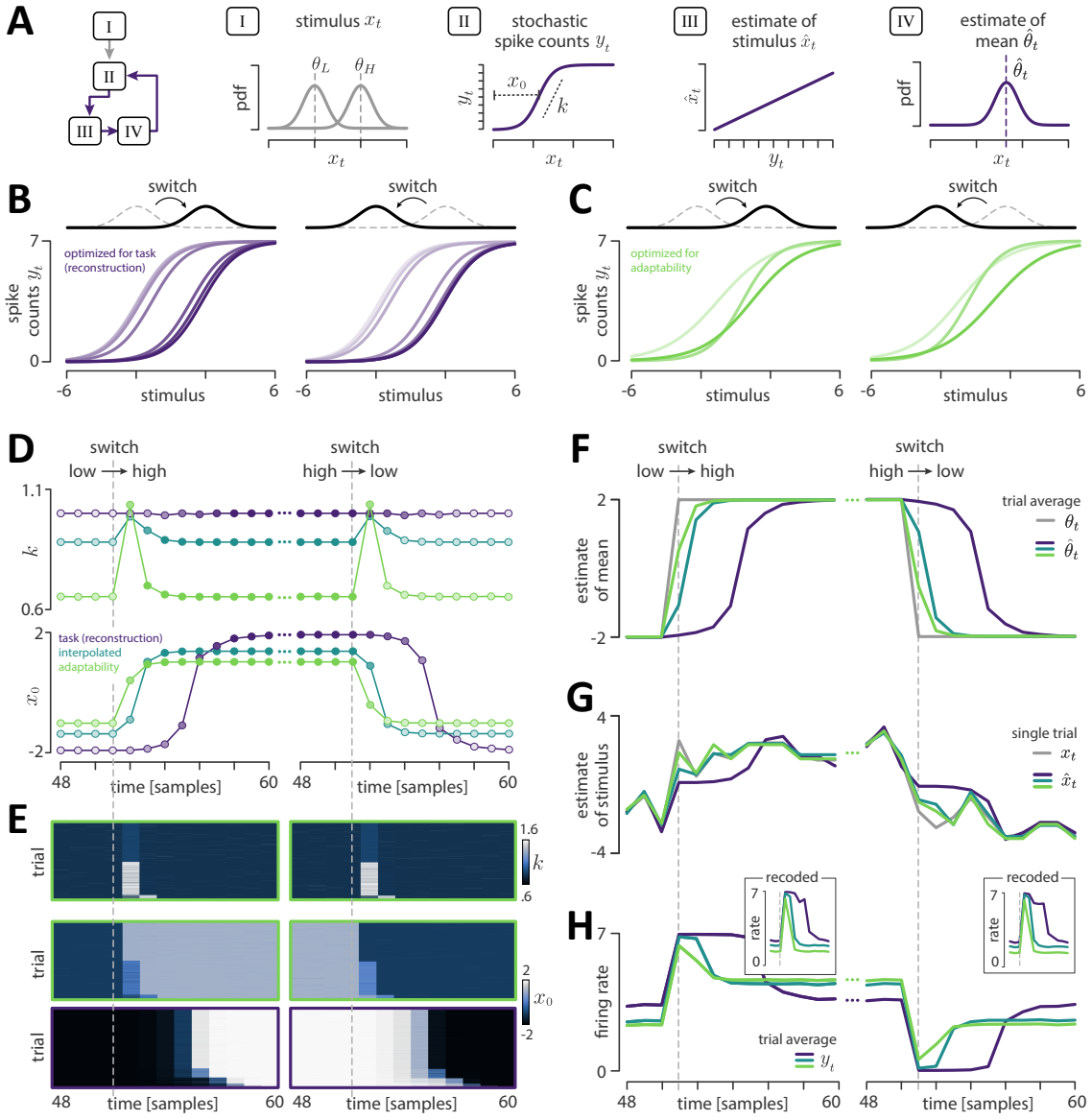

Figure 2: Neural dynamics during mean adaptation. A. Stimuli x_t are drawn from a Gaussian stimulus distribution whose variance is fixed but whose mean is determined by a latent parameter θ_t that switches stochastically between a low value, θ_L, and a high value, θ_H (we take $= -2$, $\theta_H = 2$). Stimuli are encoded in discrete spike counts y_t via a saturating nonlinearity whose slope k and offset x_0 can adapt over time. The system uses these spike counts to decode an estimate $\hat{x}_t$ of the stimulus, which is then used to update an estimate $\hat{\theta}_t$ of the mean of the incoming stimulus distribution. This estimate is used to optimally adapt the encoding nonlinearity for task performance (stimulus reconstruction) or adaptability. **B-D.** Following a switch in the stimulus mean, a nonlinearity optimized for reconstruction (B) is characterized by a constant slope (D; upper row) and an offset that tracks track the internal estimate of the stimulus mean (D; lower row). A nonlinearity optimized for adaptability (C) is characterized by a slope that transiently increases (D; upper row) and an offset that transiently shifts to align with the midpoint between the two stimulus distributions (D; lower row). An intermediate strategy (turquoise) interpolates between these extremes. Color saturation indicates time after switch. **E.** On individual trials, parameters of optimal nonlinearities show rich temporal dynamics that are not captured by their trial averages (D). A nonlinearity optimized for adaptability (green) shows two distinct behaviors. On a majority of trials, there is a brief (lasting only a single timepoint) increase in slope and shift in offset before both parameters reach their steady-state values. On the remaining trials, the parameters abruptly switch to their new steady-state values within in a single timestep. A nonlinearity optimized for reconstruction (purple) shows a more gradual change in offset that occurs at varying times following the change in stimulus mean. **F-H.** During this transient period of adaptation, differences in nonlinearity dynamics impact how quickly the system can detect changes in the stimulus mean (F), and how accurately the system can reconstruct stimuli that signal these changes (G). This adaptation is signaled by an abrupt change in firing rate following a switch (H). A simple recoding procedure (insets) generates firing rate dynamics that are symmetric to upward and downward switches in the stimulus mean. All results in panels B-D, F, H are averaged over 5000 switches in the stimulus distribution. Results in panel E are sorted by the time that the slope of the nonlinearity first reached 90% of its peak value.

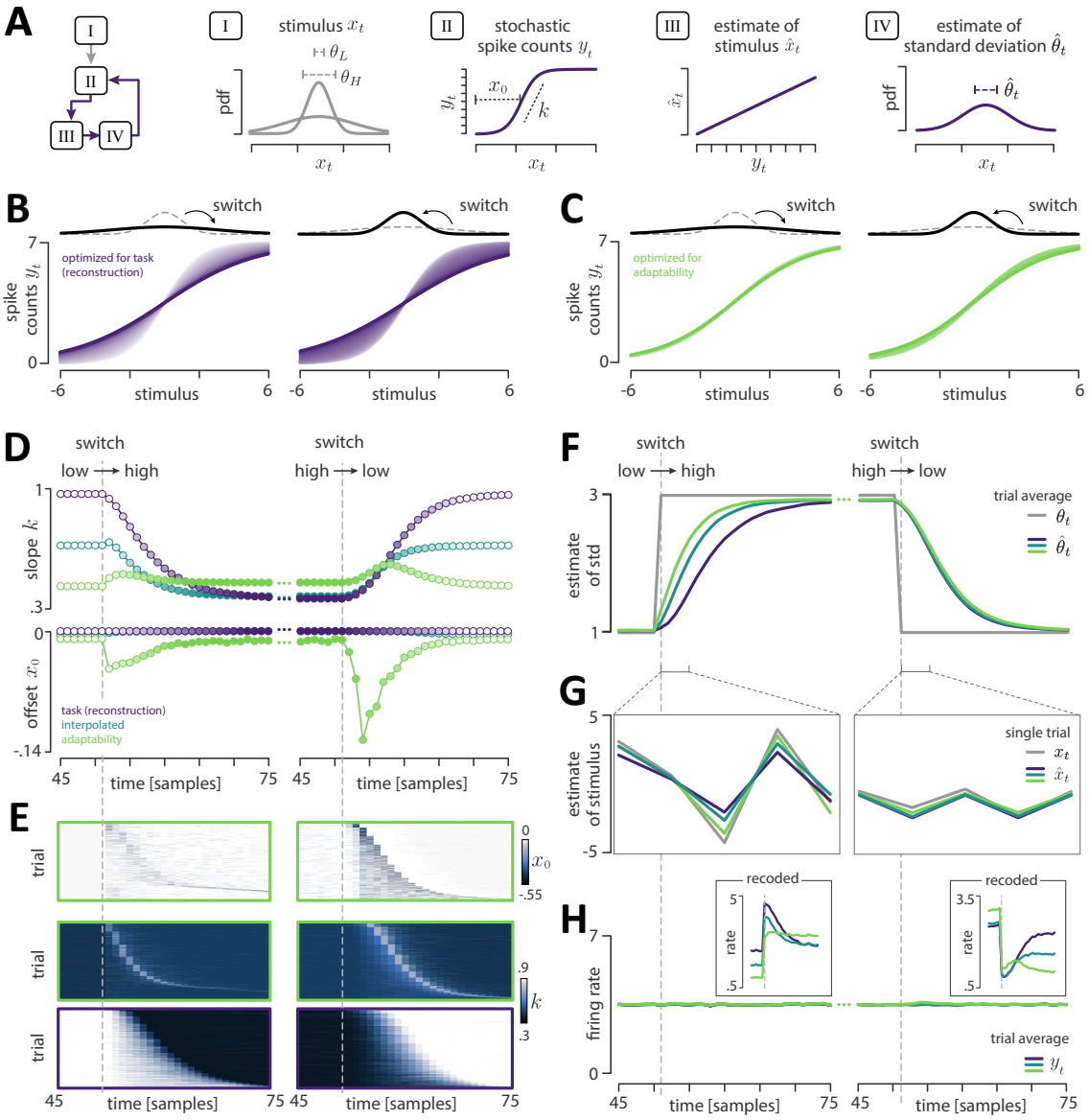

Figure 3: Neural dynamics during variance adaptation. A. Stimuli x_t are drawn from a Gaussian stimulus distribution whose mean is fixed but whose standard deviation is determined by a latent parameter θ_t that switches between θ_L and θ_H. Stimuli are encoded and decoded as described in Fig 2A, and used to update an estimate $\hat{\theta}_t$ of the standard deviation of the stimulus distribution. This estimate is used to optimally adapt the encoding nonlinearity for task performance (stimulus reconstruction) or adaptability. **B-D.** Following a switch in stimulus variance, a nonlinearity optimized for reconstruction (B) is characterized by a constant offset (D; lower row) and a slope that scales with system's internal estimate of the stimulus variance (D; upper row). A nonlinearity optimized for adaptability (C) shows a transient increase in slope (D; upper row) and decrease in offset (D; lower row) following the switch. An intermediate strategy (turquoise) interpolates between these extremes. Color saturation indicates time during the transient period of adaptation. **E.** On individual trials, parameters of optimal nonlinearities show rich temporal dynamics that are not captured by their trial averages (D). A nonlinearity optimized for reconstruction (purple) shows a more gradual change in slope that occurs at varying times following the change in stimulus variance. In comparison, a nonlinearity optimized for adaptability shows a prolonged decrease in offset followed by a brief change in slope. **F-H.** During this transient period of adaptation, differences in nonlinearity dynamics impact how quickly the system can detect changes in the stimulus variance (F), and how accurately the system can reconstruct stimuli that signal these changes (G). This adaptation is not signaled by a change in firing rate (H). A simple recoding procedure (insets) generates firing rate dynamics that show transient changes following a change in stimulus variance. In a system optimized for reconstruction, these changes track changes in stimulus variance (increasing after a switch from low to high, and decreasing after a switch from low to high; insets). In contrast, a system optimized for adaptability shows transient increases in firing rate regardless of whether stimulus variance increased or decreased. All results in panels B-D, F, H are averaged over 5000 switches in the stimulus distribution. Results in panel E are sorted by the time that the slope of the nonlinearity first reached 90% of its peak value.

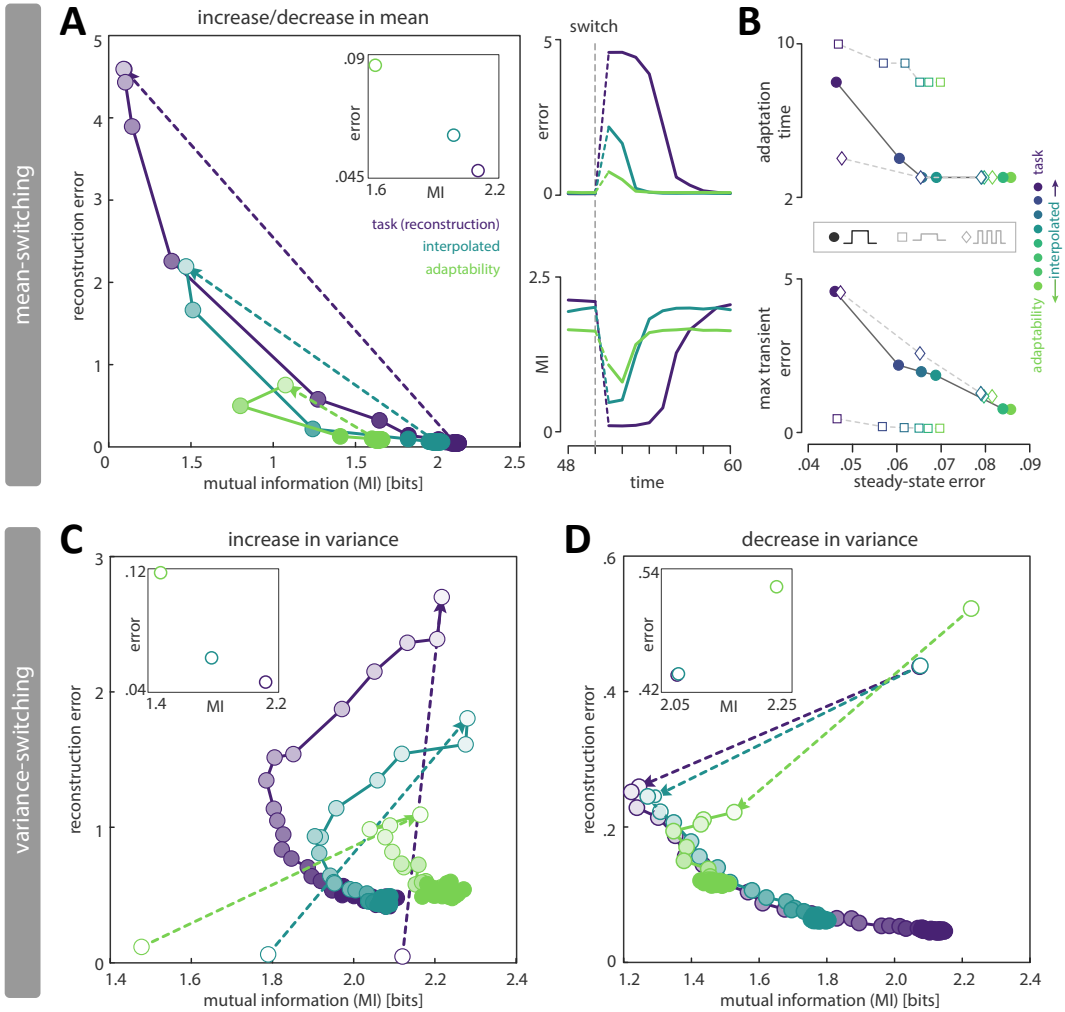

Figure 4: Rate-distortion analysis of adaptation dynamics. A. In steady state (inset), nonlinearities optimized for reconstruction (purple) maintain higher information rates and lower reconstruction error than nonlinearities optimized for adaptability (green). Following an abrupt change in the mean of the stimulus distribution (dotted arrows), both systems exhibit an abrupt increase in error and drop in information, followed by a transient period of adaptation (solid lines and filled markers). **B.** Although a task-optimized encoding maintains lower error in steady state, it reaches higher transient error and takes longer to adapt than an encoding optimized for adaptability (circular markers). Mixed codes that balance task performance and adaptability can achieve performance that interpolates between these extremes (also indicated by the turquoise trajectories in panels A, C, and D, and in Figs. 2-3). Decreasing the variability of the environment (square markers) leads to an increase in adaptation time and decrease in maximum error, while increasing the volatility of the environment (diamond markers) has the opposite effect. **C-D.** When the variance of the stimulus distribution is changing in time, the dynamics on the rate-distortion plane depend on whether the stimulus variance increases (C) or decreases (D). In both cases, the task-optimized encoding maintains lower steady-state task error (inset) but higher transient error than an encoding optimized for adaptability, and takes longer to adapt to a change in variance. Mutual information and average reconstruction error are computed over 5000 switches in the stimulus distribution.

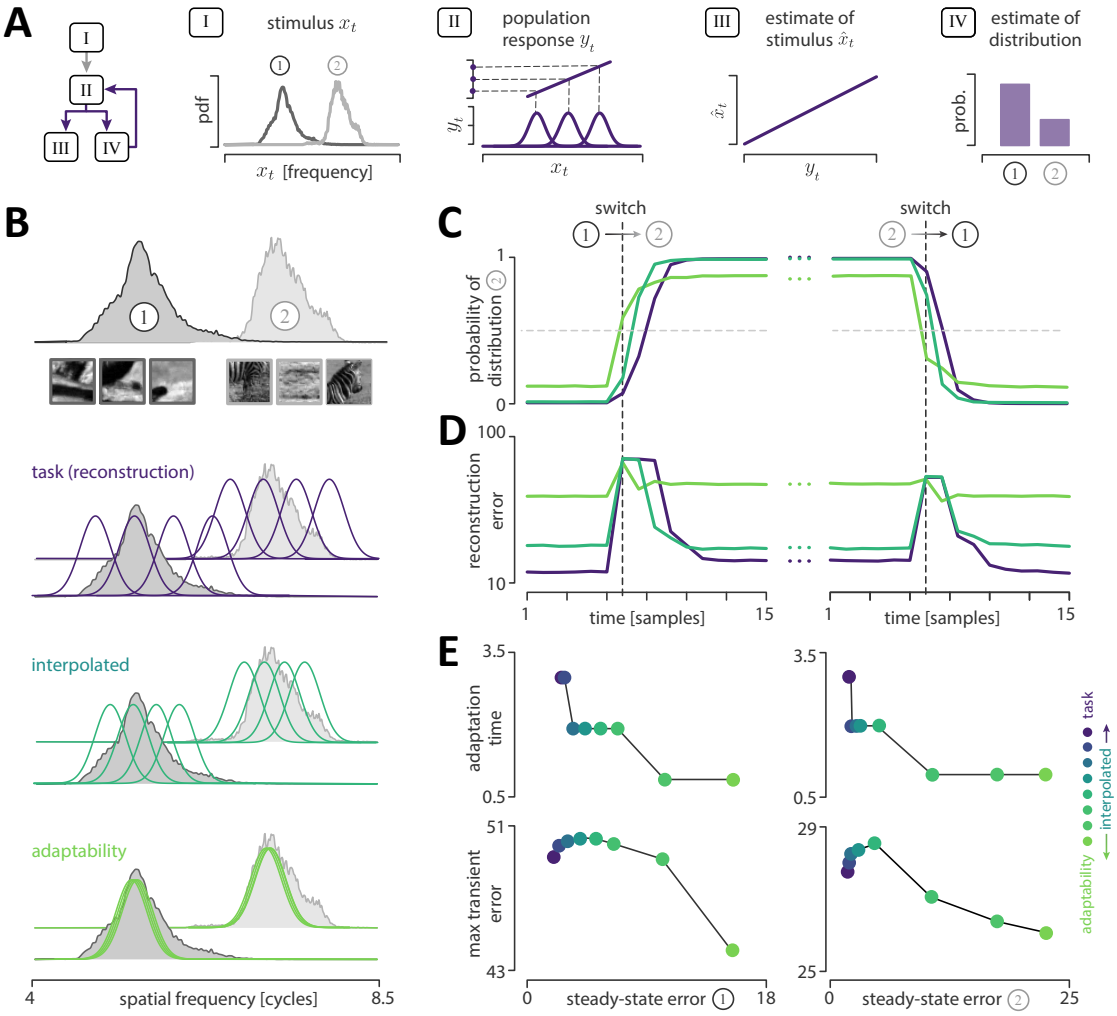

Figure 5: Dynamics of adaptation in neural populations. A. Stimuli x_t are drawn from one of two different distributions of spatial frequency, derived from snippets of video footage (I). These stimuli are encoded by a population of neurons whose response y_t is determined by Gaussian tuning curves of fixed width. The locations of these tuning curves are parametrized by a linear density function (II). The system uses the population response to linearly decode an estimate $\hat{x}_t$ of the stimulus (III). To adapt the population code, the system must infer which of the two environments was most likely to have generated incoming stimuli (IV). **B.** Populations optimized for adaptability (bottom row) maximize the difference in population responses generated by the two stimulus distributions. Their tuning curves are therefore concentrated around modes of each stimulus distribution. Populations optimized for stimulus reconstruction (top row) have tuning curves that uniformly tile the range of stimuli from each distribution. An intermediate coding strategy (middle row) produces tuning curves that interpolate between these extremes. **C-D.** Population codes optimized for reconstruction (purple), adaptability (green), or a mix of both (turquoise) respond with different latencies in detecting a switch in the stimulus distribution (panel C), and achieve different levels of performance during this transient period (panel D). **E.** The varied dynamics in panels C-D give rise to consistent tradeoffs between the steady-state reconstruction error, the time required for adaptation, and the maximum reconstruction error achieved during adaptation.

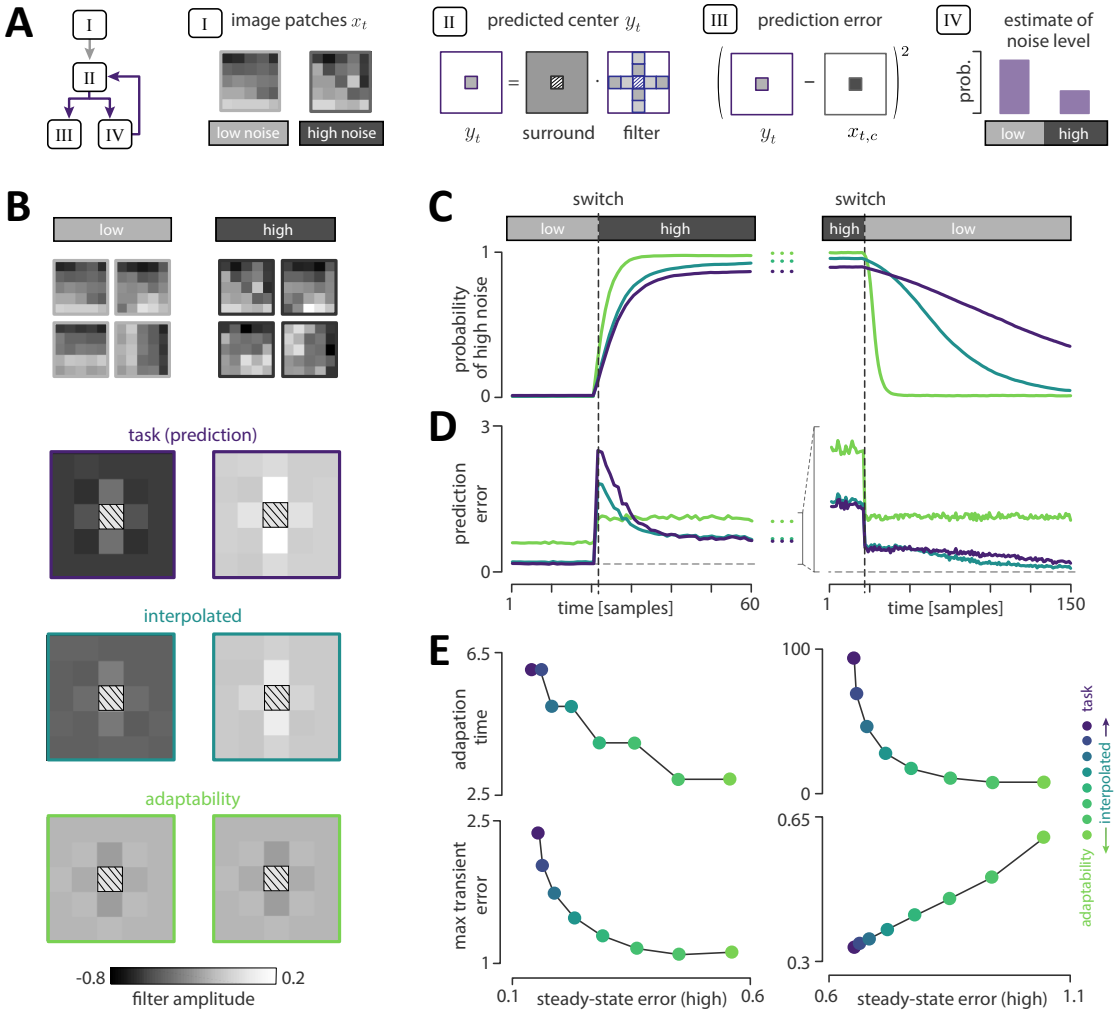

Figure 6: Dynamics of adaptive predictive coding. A. Stimuli are drawn from an ensemble of 5×5 pixel image patches. Image patches are then corrupted by either low or high levels of noise (6 dB and 20 dB SNR respectively; I). The encoder predicts the value of the center pixel y_t by linear combining the image surround with a predictive coding filter (II). The predictive filter is designed to minimize the squared error between the predicted (y_t) and actual ($c_{t,c}$) value of the central pixel, which improves coding efficiency (III). To adapt the encoding filter, the system must infer whether stimuli are corrupted by low or high levels of noise (IV). **B.** Optimal predictive coding filters depend on the level of noise in the environment (top row). A filter optimized for adaptability maximizes the difference in response statistics across the environments (bottom row). An interpolated filter is constructed by linearly combining filters optimized for prediction and adaptability (middle row). **C-D.** Filters that are optimized for prediction (purple), adaptability (green), or a mix of both (turquoise) differ in how quickly they can signal a change in noise level (C), and how much prediction error they induce during this transient period (D). **E** The varied dynamics in panels C-D give rise to tradeoffs between the steady-state reconstruction error, the time required for adaptation, and the maximum prediction error achieved during adaptation. Maintaining lower error in steady state leads to longer adaptation times, regardless of whether the noise level increases or decreases, and leads to higher transient errors when the noise level increases but lower transient error when the noise level decreases.

METHODS

Adaptation in single neurons

Model environment. We considered a nonstationary environment whose dynamics were determined by a single latent parameter θ_t that could switch between two values, θ_L and θ_H, at a fixed hazard rate $h = 0.01$. We took θ_t to parameterize a probability distribution $p(x_t|\theta_t)$ over stimuli x_t. We considered two environments; so-called 'mean-switching', in which θ_t parameterized the mean of a Gaussian stimulus distribution with fixed variance $\sigma^2 = 1$ (i.e., $p(x_t|\theta_t) = \mathcal{N}(x_t; \theta_t, 1)$), and 'variance-switching', in which θ_t parameterized the standard deviation of a Gaussian stimulus distribution with fixed mean $\mu = 0$ (i.e., $p(x_t|\theta_t) = \mathcal{N}(x_t; 0, \theta_t^2)$). In the mean-switching environment, we took θ_t to switch between $\theta_L = -2$ and $\theta_H = 2$; in the variance-switching environment, we took θ_t to switch between $\theta_L = 1$ and $\theta_H = 3$.

Model neuron. We modeled a single neuron whose response y_t to a stimulus x_t was determined by nonlinear function $f(x_t) = 1/(1 + \exp(-k(x_t - x_0)))$ whose slope k and offset x_0 were adaptable over time. We added Gaussian noise with standard deviation $\sigma^2 = 0.01$ to the output of this nonlinearity. We then thresholded values below 0 or above 1, and we discretized the output into n response levels of equal width, ranging from 0 to $n-1$. We took this discrete value (representing a spike count) to be the neural response y_t. All results were generated with $n = 8$ response levels.

Model observer. We modeled a Bayesian observer that constructs a posterior distribution over the latent parameter θ. To this end, the neural response y_t was first decoded via a linear function with an adaptable slope p_1 and offset p_0 to reconstruct a stimulus estimate $\hat{x}_t = p_1 y_t + p_0$. This estimate was then used to update the posterior distribution over θ_t. Because the latent parameter could take only two values, the posterior can be summarized by the probability that the environment is in the low state at time t, given by $P_t^L = p(\theta_t = \theta_L|\hat{x}_{\tau \le t})$. On each timestep, P_t^L was updated with incoming stimulus estimates (as derived in [70]). We averaged the posterior to compute the point estimate $\hat{\theta}_t = P_t^L \theta_L + (1 - P_t^L)\theta_H$, which could then be used to estimate the stimulus distribution $p(x_t|\hat{\theta}_t)$ (where $p(x|\hat{\theta}) = \mathcal{N}(x; \hat{\theta}, 1)$ in a mean-switching environment, and $p(x|\hat{\theta}) = \mathcal{N}(x; 0, \hat{\theta}^2)$ in a variance-switching environment). This estimate was fed back upstream and used to determine the optimal encoding and decoding parameters on the next timestep.

Optimization of nonlinear response functions. The parameters of the nonlinear encoder and the linear decoder were optimized on every timestep given the system's current posterior P_t^L. In practice, we discretized the posterior into 101 evenly spaced values between 0 and 1. For each value of the posterior, we performed an exhaustive grid search over a 201×201 grid of parameter values, spanning the ranges $[0, 2]$, $[0, 1.5]$ for the slope k (in the mean- and variance-switching environments, respectively), and $[-2.5, 2.5]$, $[-1, 0]$ for the offset x_0 (mean- and variance-switching, respectively).

For each value of the posterior (and each corresponding value of $\hat{\theta}$), we generated 100000 stimulus samples drawn from $p(x|\hat{\theta})$. We passed this set of stimulus samples through the encoding nonlinearity, defined by a given combination of slope k and offset x_0, to generate a distribution of spike counts y. We computed the parameters p_0 and p_1 of the optimal linear decoder that minimized the reconstruction error $\langle (x - \hat{x})^2 \rangle$ between the true x and reconstructed $\hat{x}$ stimulus values. The resulting set of reconstructed stimuli were used to update the estimate $\hat{\theta}(\hat{x})$ and compute the average inference error $\langle (\hat{\theta}(x) - \hat{\theta}(\hat{x}))^2 \rangle$. When iterated over all combinations of x_0 and k, this procedure resulted in two error landscapes, $E_{\text{inf}}(k, x_0|P_t^L)$ and $E_{\text{rec}}(k, x_0|P_t^L)$, for each value of the posterior P_t^L.

In the mean-switching environment, the symmetry of the problem guarantees that the error landscapes are equivalent under the exchange $x_0 \to (-x_0)$ and $P_t^L \to (1 - P_t^L)$. We used this equivalence to reduce numerical noise in our optimization through the following construction: $E(k, x_0|P_t^L) = \frac{1}{2}[E(k, x_0|P_t^L) + E(k, -x_0|1 - P_t^L)]$. This was not performed for the variance-switching environment. We then normalized the values of each landscape to lie between 0 and 1, and we smoothed each landscape by replacing the error at a given entry (x_0, k) with the average over a 12×12 region around the entry. We then constructed weighted combinations of the two landscapes: $E_\epsilon(k, x_0|P_t^L) = \epsilon E_{\text{inf}}(k, x_0|P_t^L) + (1 - \epsilon)E_{\text{rec}}(k, x_0|P_t^L)$, where ϵ is a weighting factor that we varied between 0 (pure inference) and 1 (pure reconstruction). Finally, we found the combination of parameters (x_0, k) (with corresponding decoding parameters (p_0, p_1)) that minimized the error $E_\epsilon(k, x_0|P_t^L)$ for each value of P_t^L and each value of ϵ.

In the main text, the 'adaptability-optimized' strategy was generated with $\epsilon = 0$, and the 'task-optimized' strategy was generated with $\epsilon = 1$. The interpolated strategy shown in Figs. 2, 3, and 4A,C,D was generated with $\epsilon = 0.8$.

The interpolated strategies shown in Fig. 4B were generated with $\epsilon = [0.2, 0.4, 0.6, 0.8]$. We found empirically that when optimized for reconstruction ($\epsilon = 0$), the optimal offset in the mean-switching environment tracked the system's current estimate of the mean of the stimulus distribution, and the optimal offset in the variance-switching environment scaled inversely with the system's current estimate of the standard deviation of the stimulus distribution.

Simulations. We simulated a probe environment that switched between θ_L and θ_H every 100 timesteps; this probe signal is not unlikely given the generative process for θ_t [70]. At each timestep t, the system maintained a posterior belief P^L_{t-1} and corresponding point estimate $\hat{\theta}_{t-1}$ from the previous timestep. The posterior belief specified the set of optimal encoding and decoding parameters to be used on the current timestep (as described above). We then randomly sampled a single stimulus sample x_t from $p(x_t|\theta_t)$. This stimulus sample was encoded and decoded to construct an estimate $\hat{x}_t$, which was then used to update the posterior belief [70].

We defined a single trial of the probe environment to consist of 100 timesteps in the low state ($\theta_t = \theta_L$), followed by 100 timesteps in the high state ($\theta_t = \theta_H$). We simulated dynamics over 5000 continuous trials, and we averaged the results across trials. We then separately analyzed the average response of the system to upward and downward switches in θ_t.

Computation of rate-distortion dynamics. At each timestep, we estimated the mutual information between the neuron's response y_t and the stimulus x_t from the joint histogram $p(x_t, y_t)$ measured across trials. Stimuli were partitioned into n different bins that were chosen on each timepoint to partition the empirical distribution of stimuli into bins of equal probability. This partitioning was used to estimate the marginal entropy $H(X)$. Responses were binned according to their discrete values and used to estimate the marginal entropy $H(Y)$. The joint set of bins to was then used to estimate the joint entropy $H(X, Y)$ and the mutual information $MI = H(X) + H(Y) - H(Y|X)$. We used $n = 8$ bins to partition stimuli (equal to the number of discrete response levels of the nonlinear encoder). We note that, due to finite sampling effects, the mutual information slightly exceeds the differential entropy of the stimulus x_t in the mean-switching environment.

We measured the average distortion at each time-point by computing the trial average error in reconstruction $\langle (x_t - \hat{x}_t)^2 \rangle$. Together with time-course of mutual information, this specified a dynamic trajectory on the rate-distortion plane.

Computation of tradeoffs. We defined the transient period of adaptation as the period of time following a switch in θ_t for which the relative magnitude of the time-derivative of the posterior, $|dP^L_t/dt|/\max(|dP^L_t/dt|)$, was greater than a fixed threshold a. We used $a = .025$ in the mean-switching environment, and $a = .05$ in the variance-switching environment. We defined the steady-state period as the time after adaptation and before the next switch in θ_t. We reported tradeoffs between the duration of adaptation, the maximum reconstruction error measured during adaptation, and the average reconstruction error measured during the steady-state period before the switch. In Fig. 4B, we examined the dependence of these tradeoffs on the variability and volatility of the stimulus distribution in the mean-switching environment. We varied the 'variability' of the stimulus distribution by decreasing the separation between latent states (such that $\theta_L = -1$ and $\theta_H = 1$). We varied the 'volatility' of the stimulus distribution by increasing the hazard rate to $h = 0.1$. All other parameter optimizations and simulations were carried out as described above.

Entropy coding. We used an entropy coding procedure to recode the discrete output of the encoding nonlinearity. Entropy coding assigns codewords to response levels based on the predicted probability that the response level will be used, such that shortest codewords (lowest spike counts) are assigned to most probable levels [35]. In our case, we used entropy coding to reassign spike counts to the n output response levels of the encoding nonlinearity based on the system's current prediction of the stimulus distribution. We performed this recoding for each timepoint and on each individual trial. We first used the system's internal estimate of the stimulus distribution, $p(x_t|\hat{\theta}_{t-1})$, to rank-order the response levels in terms of their expected probability of being used (i.e., by passing stimuli from $p(x_t|\hat{\theta}_{t-1})$ through the encoding nonlinearity, and then computing the histogram over response levels). We then assigned spike counts to response levels based on decreasing expected probability, such that the response level that had the highest expected probability was assigned the fewest spikes, and the response level that had the lowest expected probability was assigned the most spikes. We used this mapping to assign a spike count y_t to the stimulus sample x_t that was drawn at the given timepoint and on the given trial. This resulted in a new recoded raster of spike counts that we used to compute the trial-averaged firing rates in the insets of Figs. 2H and 3H.

Adaptation in neural populations

Model environment. We computed distributions of average spatial frequency from two different ensembles of images taken from excerpts of a natural movie [38]. The image ensembles were chosen by hand to ensure a large separation between the two distributions. From each ensemble, we randomly-sampled a set of 20000 image patches, each of size 40×40 pixels. We computed a 2-dimensional Fourier transform of each patch, and we separately averaged the power spectrum along the horizontal and vertical dimensions of the patch. We then averaged these two vectors to obtain an average power vector, and we used this vector to compute a weighted average over frequencies for each patch. This resulted in two distributions of spatial frequencies corresponding to lower and higher spatial frequencies (denoted L and H, respectively). We considered an environment that could switch between the two distributions, $F_t = L$ and $F_t = H$, at a small but fixed hazard rate $h = 0.01$, where $F_t \in \{L, H\}$ denotes the current frequency distribution at time t.

Model population. We modeled a population of N neurons whose response $\mathbf{y_t}$ to a spatial frequency x_t was determined by a set of Gaussian tuning curves. Each neuron i was described by a tuning curve with fixed standard deviation $\sigma_i = 0.2$ and adaptable peak location μ_i. We parameterized the peak locations with a linear density function [26, 39]. We first defined a set of locations μ_i^U that were uniformly distributed on the interval $[0, 1]$, and then we mapped these locations onto the stimulus axis via a linear density function: $\mu_i = a\mu_i^U + b$. In this manner, the entire population was described by only two parameters: the slope a and the offset b of the linear function.

As in our single neuron model, we mapped the output of each tuning curve onto a discrete number of response levels, and we polluted the output with additive Gaussian noise with standard deviation $\sigma = 0.01$. All results were generated using a population of $N = 4$ neurons, each with a maximum firing rate of 4 spikes (corresponding to $n = 5$ response levels).

Model observer. We derived a Bayesian observer that constructs a posterior distribution over spatial frequencies (low or high) based on the population response $\mathbf{y_t}$. As in the case of mean and variance estimation, this posterior distribution can be summarized with a single value, namely the probability that the stimuli are generated from the low-frequency distribution, given the history of past population responses: $P_t^L = p(F_t = L|\mathbf{y}_{\tau \le t})$. On each timestep, the posterior probability is updated according to the following update rule:

$$P_t^L = p(F_t = L|\mathbf{y}_{\tau \le t}) = p(\mathbf{y}_t|a, b, F_t = L)p(F_t = L|\mathbf{y}_{\tau < t})\frac{1}{p(\mathbf{y}_t)}, \tag{1}$$

where the prior probability includes the predictive component $p(F_t = L|\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1) = p(F_{t-1} = L|\mathbf{y}_{t-1}, \ldots, \mathbf{y}_1)$ $p(F_t = L|F_{t-1} = L)$. Because we consider a small hazard rate, it is reasonable for the observer to assume that environment will not change from one timestep to the next; i.e., $p(F_t = L|F_{t-1} = L) = 1$ [27, 70]. For simplicity, we also assumed that the responses of individual neurons were independent across the population:

$$p(\mathbf{y}_t|a, b, F) = \prod_{i=1}^{N} p(y_{i,t}|a, b, F), \tag{2}$$

and we estimated the conditional probabilities $p(y_{i,t}|a, b, F)$ by computing histograms of spike counts for each frequency distribution.

This posterior was used to estimate, at each timepoint, whether stimuli were more likely to be drawn from the low-frequency distribution ($P_t^L > 0.5$) or the high-frequency distribution ($P_t^L < .5$). This information was then fed back upstream and used to determine the optimal distribution of population tuning curves on the next timestep.

Optimization of tuning curves. The parameters of the linear density function were optimized separately for each frequency distribution. We performed an exhaustive search over a 32×32 grid of parameter values, spanning the range $[1, 4]$ for the slope a, and $[6, 12]$ for the offset b. For each combination of parameters on the grid, we computed the population response $\mathbf{y_F}$ to all stimulus samples drawn from a given frequency distribution $F \in \{L, H\}$. We then fit a set of linear decoding weights $\mathbf{w_F}$ that minimized the reconstruction error $\langle(\hat{x}_F - x)^2\rangle$, where $\hat{x}_F = \mathbf{w_F}\mathbf{y_F}$ was the reconstructed stimulus value. We fit these weights separately for each frequency distribution and for each combination of parameters (a, b) on the grid.

To optimize for the task of reconstruction ('R'), we determined the set of encoding parameters (a_F^R, b_F^R) (and corresponding decoding weights $\mathbf{w_F^R}$) that minimized the reconstruction error in each environment. To optimize for

adaptability ('A'), we determined the set of parameters $(a_F^A, b_F^A, \mathbf{w_F^A})$ that maximized the discriminability between the two environments. To do this, we first computed the symmetrized Kullback-Leibler divergence between the distribution of population responses to the two frequency distributions:

$$D(a,b) = \left[D_{KL}\Big(p(\mathbf{y_t}|a,b,F=L) \| p(\mathbf{y_t}|a,b,F=H)\Big) + D_{KL}\Big(p(\mathbf{y_t}|a,b,F=H) \| p(\mathbf{y_t}|a,b,F=L)\Big) \right] \quad (3)$$

For the two distributions considered here, the function $D(a,b)$ had two local maxima. Each maximum corresponded to a population where all tuning curves were concentrated around the mode of one of the two frequency distributions. For each distribution, we thus chose the set of optimal encoding parameters (a_F^A, b_F^A) (and decoding weights $\mathbf{w}_F^A$) corresponding to the mode of that distribution.

To construct interpolated population codes ('I'), we linearly combined the parameters that were optimized for reconstruction versus adaptability: $a_F^I = \epsilon a_F^A + (1-\epsilon) a_F^R$, and $b_F^I = \epsilon b_F^A + (1-\epsilon) * b_F^R$. The interpolated strategy shown in Fig. 5E was generated using $\epsilon = 0.43$. The interpolated strategies shown in Fig. 5E were generated using 8 values of ϵ spaced uniformly between 0 and 1.

Simulations. As before, we simulated a probe environment that switches between the two stimulus distributions, $F_t = L$ and $F_t = H$, every 100 timesteps. At each timestep t, the system maintained a posterior belief P_{t-1}^L about the current stimulus distribution. This belief specified the optimal set of parameters (a_F, b_F) that defined the population code. When $P_{t-1}^L < 0.5$, the system believed that stimuli were more likely to be coming from the high-frequency distribution, and thus the system used a population code parameterized by (a_H, b_H) and a decoder parameterized by ϕ_H. When $P_{t-1}^L > 0.5$, the system used a population code and decoder parameterized by (a_L, b_L, ϕ_L). A single stimulus sample x_t was then randomly drawn from the stimulus distribution defined by F_t. The sample was encoded and decoded as described above, and used to update the posterior.

We defined a single trial to consist of 100 timesteps in which stimuli were randomly drawn from the low-frequency distribution ($F_t = L$), followed by 100 timesteps in which stimuli were drawn from the high-frequency distribution ($F_t = H$). We simulated dynamics over 10000 continuous trials, and we averaged the results across trials.

Computation of tradeoffs. We defined the transient period of adaptation as the period of time following a switch in F_t for which the the posterior, P_t^L, remained above (switch from low to high frequency distribution) or below (switch from high to low frequency distribution) a threshold of 0.5. This hard thresholding reflects the discrete nature of the inference problem, and the fact that the encoding changes once the posterior crosses the 0.5 threshold. We defined the steady-state period as the first set of 50 samples after adaptation and before the next switch in F_t.

Adaptation of predictive filters

Model environment. We sampled 20000 image patches, each of size 5×5 pixels, from the low-frequency image ensemble described in the previous section. We subtracted the mean pixel value from each patch, and divided each pixel in the patch by the root-mean-squared (RMS) value computed across pixels. To simulate different noise conditions, we distorted each patch with additive Gaussian noise whose variance was either low ($\sigma = 0.1$, corresponding to an average signal to noise ratio (SNR) of 20 dB), or high ($\sigma = 0.5$, corresponding to an SNR of 6 dB). We considered an environment that could switch between the two noise conditions, $N_t = L$ (low noise) and $N_t = H$ (high noise), at a hazard rate $h = 0.001$, where $N_t \in \{L, H\}$ denotes the noise level at time t.

Model neuron. We modeled a single neuron whose response y_t to an image patch $\mathbf{x_t}$ was determined by a adaptable filter ϕ_N (with $N \in \{L, H\}$). Specifically, we took y_t to be the dot product between ϕ_N and the surround $\mathbf{x}_{s,t}$ of the image patch: $y_t = \phi_N \mathbf{x}_{s,t}$.

Model observer. We derived a Bayesian observer that constructs a posterior distribution over noise levels (low or high) based on the filter output y_t. Again, this posterior distribution can be written as the probability that the noise level is low, given the history of past filter outputs: $P_t^L = p(N_t = L | y_{\tau \le t})$. The posterior has same the form as given in Eq. 1, and we again assume $p(N_t = L | N_{t-1} = L) = 1$. We also assumed that the distribution of filter outputs y

in each of the noise conditions was Gaussian with a mean μ_N and variance σ_N^2 that depended on the noise level: $p(y_t|N) = \mathcal{N}(\mu_N, \sigma_N^2)$.

As before, this posterior was used to estimate whether the the noise level was low ($P_t^L > 0.5$) or high ($P_t^L < .5$), and this information was then fed back upstream and used to determine the optimal filter on the next timestep.

Optimization of predictive filters. To optimize for the task of prediction ('P'), we learned separate filters ϕ_N^P for the low and high noise conditions. We used gradient descent to minimize the prediction error $\langle (x_{c,t} - y_t)^2 \rangle$ between the central pixel $x_{c,t}$ and a weighted average of the surround, $y_t = \phi_N^P \mathbf{x}_{s,t}$.

To optimize for adaptability ('A'), we learned a single filter ϕ^A to maximize the difference of filter output distributions between low and high noise levels. This filter was used in both noise conditions, such that $\phi_L^A = \phi_H^A = \phi^A$. To learn the optimal filter, we computed the symmetrized Kullback-Leibler divergence between the output distributions in the two noise conditions:

$$D(\phi) = \left[D_{KL}\Big(p(y|\phi, \mu_L, \sigma_L^2) \| p(y|\phi, \mu_H, \sigma_H^2)\Big) + D_{KL}\Big(p(y|\phi, \mu_H, \sigma_H^2) \| p(y|\phi, \mu_L, \sigma_L^2)\Big) \right] \tag{4}$$

Optimizations were performed using the Matlab function `fminunc()`.

To construct interpolated filters ('I'), we linearly combined the filters that were optimized for prediction versus adaptability: $\phi_N^I = \epsilon \phi_N^A + (1 - \epsilon)\phi_N^P$. The interpolated strategy shown in Fig. 6B-D was generated using $\epsilon = 0.29$. The interpolated strategies shown in Fig. 6E were generated using 8 values of ϵ spaced uniformly between 0 and 1.

Simulations. As before, we simulated a probe environment that switched between the two noise levels, $N_t = L$ and $N_t = H$, every 100 timesteps. At each timestep t, the system maintained a posterior belief P_{t-1}^L about the noise level that specified the optimal filter ϕ. When $P_{t-1}^L < 0.5$, the system believed that the noise level was high and thus used the filter ϕ_H. When $P_{t-1}^1 > 0.5$, the system used the filter ϕ_L. A single image patch x_t was then drawn from the stimulus distribution, distorted with noise whose variance was determined by N_t, encoded as described above, and used to update the posterior.

We defined a single trial to consist of 1000 timesteps in which the noise level was low ($N_t = L$), followed by 1000 timesteps in which the noise level was high ($N_t = H$). We simulated dynamics over 10000 continuous trials, and we averaged the results across trials.

Computation of tradeoffs. We defined the transient period of adaptation as the period of time following a switch in N_t for which the the posterior, P_t^L, remained above (switch from low to high noise) or below (switch from high to low noise) a threshold of 0.5. As in the previous section, this hard thresholding reflects the discrete nature of the inference problem, and the fact that the encoding changes once the posterior crosses the 0.5 threshold. We defined the steady-state period as the set of 50 samples after adaptation and before the next switch in N_t.

References

1. Bergan, J. F., Ro, P., Ro, D. & Knudsen, E. I. Hunting increases adaptive auditory map plasticity in adult barn owls. *Journal of Neuroscience* **25,** 9816–9820 (2005).
2. Witten, I. B., Bergan, J. F. & Knudsen, E. I. Dynamic shifts in the owl's auditory space map predict moving sound location. *Nature neuroscience* **9,** 1439 (2006).
3. Mischiati, M. *et al.* Internal models direct dragonfly interception steering. *Nature* **517,** 333–338 (2015).
4. Thesen, A., Steen, J. B. & Doving, K. Behaviour of dogs during olfactory tracking. *Journal of Experimental Biology* **180,** 247–251 (1993).
5. Nemenman, I. 4. information theory and adaptation. *Quantitative Biology: From Molecular to Cellular Systems,* 73 (2012).
6. Fairhall, A. L., Lewen, G. D., Bialek, W. & van Steveninck, R. R.d. R. Efficiency and ambiguity in an adaptive neural code. *Nature* **412,** 787 (2001).
7. Dean, I., Harper, N. S. & McAlpine, D. Neural population coding of sound level adapts to stimulus statistics. *Nature neuroscience* **8,** 1684 (2005).
8. Clemens, J., Ozeri-Engelhard, N. & Murthy, M. Fast intensity adaptation enhances the encoding of sound in Drosophila. *Nature communications* **9,** 134 (2018).
9. Gauthier, J. L. & Tank, D. W. A dedicated population for reward coding in the hippocampus. *Neuron* **99,** 179–193 (2018).
10. Boccara, C. N., Nardin, M., Stella, F., O'Neill, J. & Csicsvari, J. The entorhinal cognitive map is attracted to goals. *Science* **363,** 1443–1447 (2019).
11. Brenner, N., Bialek, W. & Van Steveninck, R. d. R. Adaptive rescaling maximizes information transmission. *Neuron* **26,** 695–702 (2000).
12. Sharpee, T. O. *et al.* Adaptive filtering enhances information transmission in visual cortex. *Nature* **439,** 936 (2006).
13. Gutnisky, D. A. & Dragoi, V. Adaptive coding of visual information in neural populations. *Nature* **452,** 220 (2008).
14. Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W. & Meister, M. Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386,** 69 (1997).
15. Schwartz, G., Harris, R., Shrom, D. & Berry II, M. J. Detection and prediction of periodic patterns by the retina. *Nature neuroscience* **10,** 552 (2007).
16. Nagel, K. I. & Doupe, A. J. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron* **51,** 845–859 (2006).
17. Ulanovsky, N., Las, L., Farkas, D. & Nelken, I. Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience* **24,** 10440–10453 (2004).
18. J, C. Multiple sites of adaptation lead to contrast encoding in the Drosophila olfactory system. *Physiological Reports* **4,** e12762 (2016).
19. Cao, L.-H. *et al.* Odor-evoked inhibition of olfactory sensory neurons drives olfactory perception in Drosophila. *Nature Communications* **8,** 1357 (2017).
20. Srinivas, G.-S., Mahmut, D., Junjianjia, L., A, C. D. & Thierry, E. Olfactory receptor neurons use gain control and complementary kinetics to encode intermittent odorant stimuli. *eLife* **6,** e27670 (2017).
21. Maravall, M., Petersen, R. S., Fairhall, A. L., Arabzadeh, E. & Diamond, M. E. Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS biology* **5,** e19 (2007).
22. Clarke, S. E., Longtin, A. & Maler, L. Contrast coding in the electrosensory system: parallels with visual computation. *Nature Reviews Neuroscience* **16,** 733 (2015).
23. Barlow, H. B. *et al.* Possible principles underlying the transformation of sensory messages. *Sensory communication* **1,** 217–234 (1961).
24. Tishby, N., Pereira, F. C. & Bialek, W. The information bottleneck method. *arXiv preprint physics/0004057* (2000).
25. Palmer, S. E., Marre, O., Berry, M. J. & Bialek, W. Predictive information in a sensory population. *Proceedings of the National Academy of Sciences* **112,** 6908–6913 (2015).
26. Wang, Z., Stocker, A. A. & Lee, D. D. Efficient neural codes that minimize lp reconstruction error. *Neural computation* **28,** 2656–2686 (2016).
27. Wark, B., Lundstrom, B. N. & Fairhall, A. Sensory adaptation. *Current opinion in neurobiology* **17,** 423–429 (2007).
28. Laughlin, S. A simple coding procedure enhances a neuron's information capacity. *Zeitschrift für Naturforschung c* **36,** 910–912 (1981).
29. Gjorgjieva, J., Sompolinsky, H. & Meister, M. Benefits of pathway splitting in sensory coding. *Journal of Neuroscience* **34,** 12127–12144 (2014).
30. Shannon, C. E., Weaver, W. & Burks, A. W. The mathematical theory of communication (1951).
31. Andrews, B. W. & Iglesias, P. A. An information-theoretic characterization of the optimal gradient sensing response of cells. *PLoS computational biology* **3,** e153 (2007).
32. Wark, B., Fairhall, A. & Rieke, F. Timescales of inference in visual adaptation. *Neuron* **61,** 750–761 (2009).
33. Baccus, S. A. & Meister, M. Fast and slow contrast adaptation in retinal circuitry. *Neuron* **36,** 909–919 (2002).

34. Rabinowitz, N. C., Willmore, B. D., Schnupp, J. W. & King, A. J. Contrast gain control in auditory cortex. *Neuron* **70,** 1178–1191 (2011).
35. Cover, T. M. & Thomas, J. A. *Elements of information theory* (John Wiley & Sons, 2012).
36. Kastner, D. B. & Baccus, S. A. Coordinated dynamic encoding in the retina using opposing forms of plasticity. *Nature neuroscience* **14,** 1317 (2011).
37. Kohn, A. & Movshon, J. A. Adaptation changes the direction tuning of macaque MT neurons. *Nature neuroscience* **7,** 764 (2004).
38. Cadieu, C. F. & Olshausen, B. A. Learning intermediate-level representations of form and motion from natural movies. *Neural computation* **24,** 827–866 (2012).
39. Ganguli, D. & Simoncelli, E. P. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural computation* **26,** 2103–2134 (2014).
40. Nirenberg, S., Bomash, I., Pillow, J. W. & Victor, J. D. Heterogeneous response dynamics in retinal ganglion cells: the interplay of predictive coding and adaptation. *American Journal of Physiology-Heart and Circulatory Physiology* (2010).
41. Hosoya, T., Baccus, S. A. & Meister, M. Dynamic predictive coding by the retina. *Nature* **436,** 71 (2005).
42. Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive coding: a fresh view of inhibition in the retina. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **216,** 427–459 (1982).
43. Kastner, D. B. & Baccus, S. A. Insights from the retina into the diverse and general computations of adaptation, detection, and prediction. *Current opinion in neurobiology* **25,** 63–69 (2014).
44. Geffen, M. N., De Vries, S. E. & Meister, M. Retinal ganglion cells can rapidly change polarity from Off to On. *PLoS biology* **5,** e65 (2007).
45. Yeh, T., Lee, B. B. & Kremers, J. The time course of adaptation in macaque retinal ganglion cells. *Vision research* **36,** 913–931 (1996).
46. Wen, B., Wang, G. I., Dean, I. & Delgutte, B. Dynamic range adaptation to sound level statistics in the auditory nerve. *Journal of Neuroscience* **29,** 13797–13808 (2009).
47. Jadzinsky, P. D. & Baccus, S. A. Synchronized amplification of local information transmission by peripheral retinal input. *Elife* **4,** e09266 (2015).
48. Drew, P. J. & Abbott, L. F. Models and properties of power-law adaptation in neural systems. *Journal of neurophysiology* (2006).
49. Lundstrom, B. N., Higgs, M. H., Spain, W. J. & Fairhall, A. L. Fractional differentiation by neocortical pyramidal neurons. *Nature neuroscience* **11,** 1335 (2008).
50. Kastner, D. B. & Baccus, S. A. Spatial segregation of adaptation and predictive sensitization in retinal ganglion cells. *Neuron* **79,** 541–554 (2013).
51. Denève, S., Alemi, A. & Bourdoukan, R. The brain as an efficient and robust adaptive learner. *Neuron* **94,** 969–977 (2017).
52. Safran, M. N., Flanagin, V. L., Borst, A. & Sompolinsky, H. Adaptation and information transmission in fly motion detection. *Journal of neurophysiology* **98,** 3309–3320 (2007).
53. Rao, R. P. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience* **2,** 79 (1999).
54. Młynarski, W. F. & Hermundstad, A. M. Adaptive coding for dynamic sensory inference. *Elife* **7,** e32055 (2018).
55. Ozuysal, Y. & Baccus, S. A. Linking the computational structure of variance adaptation to biophysical mechanisms. *Neuron* **73,** 1002–1015 (2012).
56. Ujfalussy, B. B., Makara, J. K., Branco, T. & Lengyel, M. Dendritic nonlinearities are tuned for efficient spike-based computations in cortical circuits. *Elife* **4,** e10056 (2015).
57. Chalk, M., Marre, O. & Tkačik, G. Toward a unified theory of efficient, predictive, and sparse coding. *Proceedings of the National Academy of Sciences* **115,** 186–191 (2018).
58. Shwartz-Ziv, R. & Tishby, N. Opening the black box of deep neural networks via information. *arXiv:1703.00810* (2017).
59. Lan, G., Sartori, P., Neumann, S., Sourjik, V. & Tu, Y. The energy–speed–accuracy trade-off in sensory adaptation. *Nature physics* **8,** 422 (2012).
60. Lahiri, S., Sohl-Dickstein, J. & Ganguli, S. A universal tradeoff between power, precision and speed in physical communication. *arXiv preprint arXiv:1603.07758* (2016).
61. Wei, X.-X. & Stocker, A. A. A Bayesian observer model constrained by efficient coding can explain'anti-Bayesian'percepts. *Nature neuroscience* **18,** 1509 (2015).
62. Orbán, G., Berkes, P., Fiser, J. & Lengyel, M. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* **92,** 530–543 (2016).
63. Bialek, W., Van Steveninck, R. R.D. R. & Tishby, N. *Efficient representation as a design principle for neural coding and computation* in *2006 IEEE international symposium on information theory* (2006), 659–663.
64. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381,** 607 (1996).
65. Van Hateren, J. H. A theory of maximizing sensory information. *Biological cybernetics* **68,** 23–29 (1992).

66. Krishnamurthy, K., Wark, B., Fairhall, A. & Pillow, J. *Efficient coding with time-varying stimuli and noise* in *Computational and Systems Neuroscience COSYNE* (2016).
67. Goris, R. L., Movshon, J. A. & Simoncelli, E. P. Partitioning neuronal variability. *Nature neuroscience* **17,** 858 (2014).
68. Lesica, N. A., Boloori, A. S. & Stanley, G. B. Adaptive encoding in the visual pathway. *Network: Computation in Neural Systems* **14,** 119–135 (2003).
69. Marzen, S. E. & Crutchfield, J. P. Predictive rate-distortion for infinite-order Markov processes. *Journal of Statistical Physics* **163,** 1312–1338 (2016).
70. DeWeese, M. & Zador, A. Asymmetric dynamics in optimal variance adaptation. *Neural Computation* **10,** 1179–1202 (1998).