

Feature Review

Bayesian efficient coding as a theory of perception: progress, controversies, and prospects

Ivan Tomić ^{1,*}, Máté Lengyel^{2,3}, and Paul Bays⁴

Bayesian efficient coding unifies two foundational theories of sensory processing: efficient coding and Bayesian inference. Central to this account is the idea that natural environmental statistics shape both how sensory information is encoded and how it is perceptually interpreted. By unifying these principles, the framework accounts for counterintuitive perceptual biases and establishes lawful relationships between environmental statistics, bias, and discrimination thresholds. In this article, we review behavioural and neural evidence for this theory in perception and cognition, as well as how short- and long-term adaptation to the environment may be expressed within the framework. We further review theoretical developments that extend the original framework, focusing on how response biases can be decomposed into encoding- and decoding-related components. A decade after its introduction, Bayesian efficient coding continues to evolve as a powerful theory, with recent extensions addressing early limitations and opening new directions for investigating perception and cognition.

Highlights

A growing number of studies have identified signatures of Bayesian efficient coding in perception and higher cognitive processes.

These effects not only reflect long-term environmental regularities but also update via short-term adaptation to the statistics of recent input.

Computational theory and models have revealed how perceptual biases can be decomposed into distinct encoding- and decoding-related components.

The framework has been extended with task-specific loss functions to account for behavioural phenomena not captured by the original theory.

Unifying encoding and inference in perception

To guide behaviour and support interaction with the environment, the brain encodes sensory information in neural activity and then uses it to make inferences about the state of the external world. These two components—encoding and inference (Figure 1A)—are fundamental to accounts of perception, decision-making, and learning. Classically, encoding has been investigated through the lens of **efficient coding** (see Glossary) [2–4] (Box 1), which concerns how stimuli are optimally represented in sensory responses given neural constraints, whereas **Bayesian inference** [32–36] (Box 2) has been used to explain how sensory input is combined with prior knowledge to make statistically optimal judgements. Historically, these have been treated as separate theoretical perspectives, and thus necessarily remain incomplete.

In this review, we present a general framework emerging from recent work that proposes a unification of these two theories [45–48], which we refer to as Bayesian efficient coding. According to this account, the sensory responses shaped by efficient coding constitute the inputs to Bayesian inference, and both encoding and inference are jointly constrained by the statistical structure of the sensory environment. Since its introduction, this unified framework has inspired new investigations across a wide range of perceptual phenomena, and the theory has, in turn, been expanded and modified in response to both theoretical and empirical challenges. Here, we review psychophysical, neural, and computational evidence to evaluate the strengths and limitations of Bayesian efficient coding as a theory of perception.

¹Department of Psychology, Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb, Croatia

²Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge, UK

³Center for Cognitive Computation, Department of Cognitive Science, Central European University, Budapest, Hungary

⁴Department of Psychology, University of Cambridge, Cambridge, UK

*Correspondence:
itomic3@ffzg.unizg.hr (I. Tomić).



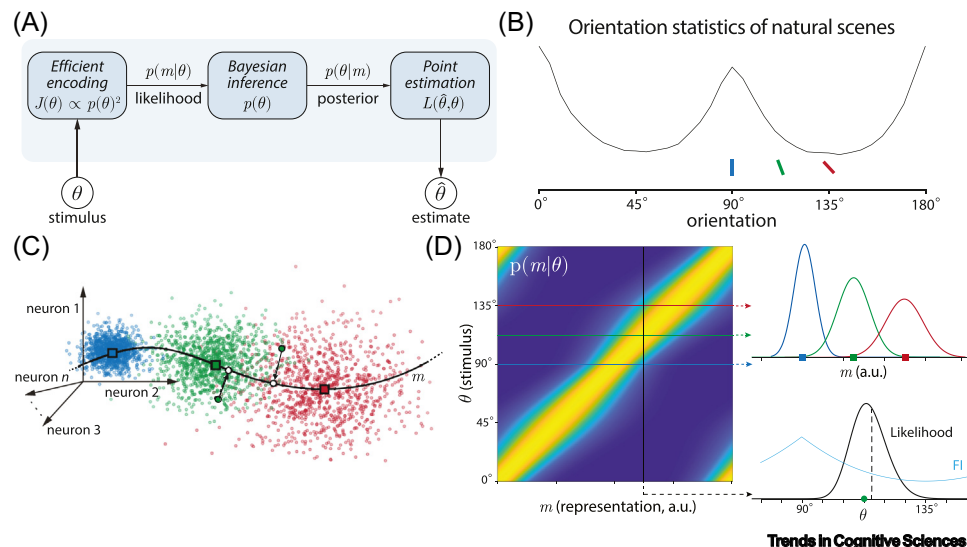


Figure 1. The Bayesian efficient coding model of perception. (A) The Bayesian efficient coding account describes perception as a probabilistic process comprising sequential encoding and decoding stages. A stimulus θ is efficiently encoded as an internal representation, m , whose likelihood is then combined with prior knowledge via Bayes' rule to compute a posterior $p(\theta|m)$. A point estimate $\hat{\theta}$ is obtained by minimising the expectation of a loss function $L(\hat{\theta}, \theta)$. (B) Distribution of visual orientations in natural images, showing the predominance of cardinal over oblique orientations. Coloured bars indicate three example orientations (90° , 112.5° , and 135°) that differ in frequency. Adapted from [1]. (C) Stimuli are encoded in a high-dimensional space of neural responses. Coloured dots illustrate noisy neural responses to the stimuli in B and coloured squares show the mean response to each. Mean responses to all stimuli lie on the black curve. In this example, any individual response (coloured discs) can be identified with the mean response corresponding to the stimulus that makes it most probable (arrows and white discs), collapsing the high-dimensional response into a one-dimensional internal representation, m . (D) The conditional probability of m given a stimulus θ (horizontal cross-sections of heat map) reflects the efficient allocation of Fisher information ($J(\theta)$; light blue curve, bottom right). Narrower distributions of m (top right) correspond to higher encoding precision. As a result, the likelihood for a given internal representation m is asymmetric (vertical cross-sections of heat map; black curve bottom right).

Bayesian efficient coding

Following Wei and Stocker [46], we present the theory based on a scalar **stimulus value** θ . In general, sensory responses to a stimulus are high-dimensional (involving the responses of many sensory neurons), but for simplicity, we assume these responses are effectively summarised by a scalar internal representation m (or 'measurement' [46]), which condenses the high-dimensional neural response (Figure 1C) into a single value while retaining all the information relevant to θ . This makes m a 'sufficient statistic'; in many cases, the maximum likelihood estimate could serve as the measurement. Figures 1 and 2 illustrate the principles of encoding and decoding, respectively, using the example of visual orientation, where the greater prevalence of cardinal (horizontal and vertical) versus oblique (diagonal) orientations in natural and built environments is well documented (Figure 1B). A key concept in formalising how resources are allocated in the process of encoding stimuli, θ , into internal representations, m , is the **Fisher information (FI)** ($J(\theta)$), which quantifies the precision with which a given stimulus value θ is represented internally.

Central to the synthesis of efficient coding and Bayesian inference is the principle that an efficient code dedicates more representational resources to stimuli that are more likely to be encountered in the environment—those with higher probability under the **prior** $p(\theta)$ —ensuring they are encoded with greater fidelity (Figure 1D). The theory further assumes that total encoding precision

Glossary

Bayes' rule: a mathematical rule for inverting conditional probabilities. The rule states that the posterior probability, $p(S|R)$, is the normalised product of the likelihood, $p(R|S)$, and prior, $p(S)$: $p(S|R) = p(R|S)p(S)/p(R)$.

Bayesian inference: computing a posterior distribution over an unknown stimulus value S using Bayes' rule. Bayesian estimation further involves selecting a single stimulus value by combining the posterior with a loss function.

Discriminability: a term sometimes used interchangeably with discrimination threshold but formally one is the inverse of the other.

Discrimination task: a psychophysical task in which an observer judges in which direction two stimuli differ along a specific, predefined dimension (e.g., direction of tilt: clockwise or counter-clockwise).

Discrimination threshold: or just-noticeable difference; the magnitude of deviation between two stimulus features that can be correctly discriminated with some benchmark probability (e.g., 90% correct).

Efficient coding: encoding of stimulus features that maximises some measure of information (e.g., mutual or Fisher information) between these features and the sensory responses they evoke, typically depending on the prevalence of these features in the environment.

Environmental statistics: the joint probability distribution of all sensory features in an individual's typical surroundings.

Estimation task: or method of adjustment; a psychophysical task in which an observer selects or adjusts a comparison stimulus on a continuous scale to match a reference stimulus.

Fisher information: a statistical measure of precision which quantifies the fidelity with which a stimulus value is encoded in sensory responses.

Image-computable model: a model that accepts an array of pixels (or other elementary stimulus features) as input and produces quantitative predictions that can be compared with neural or behavioural data.

Likelihood function: the conditional probability $p(R|S)$ evaluated as a function of S , that is, the probability with which different stimulus values S could have evoked a given sensory input R .

Box 1. Theoretical background of efficient coding

The efficient coding (EC) hypothesis [2–4] posits that sensory systems have evolved to maximise the information transmitted about the environment, while operating under biological constraints such as metabolic cost, internal noise, and limited neural resources (Wang, Z. *et al.*, 2016, NeurlPS, conference). Metabolic cost can be operationalised, for example, as the energetic cost of firing an action potential [5–7]. A neural code optimised under such a constraint tends to reduce redundancy in the transmitted signal with comparatively little loss of information.

Formally, the EC hypothesis draws on Shannon's information theory [8,9], assuming that neural responses maximise the mutual information (MI) $I(R;S)$ between stimuli S and responses R :

$$I(R;S) = H(S) - H(S|R) \quad [1]$$

where $H(S)$ is the entropy of the stimulus distribution and $H(S|R)$ is the conditional entropy of the stimulus given the response, with the latter quantifying the uncertainty about the stimulus that remains after observing the sensory response [10]. FI can be used to specify a bound on MI [11] and serves as a practical approximation to MI in the low-noise regime [12].

In practice, maximising MI requires structuring neural responses to exploit statistical regularities in the environment, so that they convey as much information as possible about the stimulus (Figures 1C,D and 2C,D). Consistent with this principle, sensory and motor systems spanning vision, audition, olfaction, somatosensation, and motor planning adapt to the distribution and dynamics of natural inputs [4, 13–19]. Experimental work has also shown that sensory neurons adjust their activity to reflect the statistics of the input signal in order to increase encoding efficiency [20–22].

Although originally developed to explain early sensory encoding in relatively stable environments (i.e., those to which sensory systems have adapted through evolution), the EC hypothesis has been applied more broadly. For example, recent work has explored the dynamics of efficient codes in nonstationary environments [23,24]. Beyond sensory processing, the EC principles have been applied to a range of cognitive and behavioural phenomena, including decision-making [25], economic behaviour under risk [26], and reward learning [27]. Finally, the core EC idea—that neural representations reflect the solution to a constrained optimisation problem—has also been formalised within theories of optimal compression (i.e., rate-distortion theory [28,29]) and has been linked to predictive coding frameworks [30,31].

is limited, such that allocating more resources to one region of the stimulus space necessarily leaves fewer resources elsewhere (see the section on 'Theoretical advances' for more detail). Under relatively mild assumptions, it can be shown mathematically (Box 3) that such an efficient

Box 2. Theoretical background of Bayesian inference

While efficient coding is concerned with how stimuli are efficiently represented in the brain (Box 1), Bayesian inference focuses on inferring latent properties of the world—such as the stimulus that generated a given sensory input [32–36]. According to this framework, perception is a probabilistic interpretation shaped by both the way sensory inputs depend on stimuli and the observer's prior knowledge about the environment (Figure 2). Formally, this inference is described by Bayes' rule, which defines the posterior probability of a stimulus S given a sensory input R in terms of the likelihood $p(R|S)$ and the prior probability of the stimulus $p(S)$:

$$p(S|R) \propto p(R|S)p(S) \quad [1]$$

This posterior distribution expresses the observer's belief about the stimulus given the current sensory evidence. Making perceptual decisions often requires selecting a single—or point—estimate from the full posterior. This choice is determined by a loss function that describes the behavioural costs of different estimation errors [37,38], for example, if the cost increases with the square of the deviation from the true value, the posterior mean provides the optimal point estimate that minimises the expected cost. Bayesian inference provides a normative account of statistically optimal perception: when sensory information is unreliable, prior knowledge should exert greater influence, biasing estimates towards more probable stimuli (Figure 2B,D). In perception, these priors are thought to reflect environmental statistics, serving as a rational basis for forming beliefs about the probable occurrence of stimuli. Priors can be shaped by evolution as well as by the observer's accumulated experience through adaptation and learning (e.g., [39–41]). Substantial evidence supports the idea that both humans and other species rely on Bayesian inference to guide perception and action (e.g., [42–44]).

While, in general, S and R can be high-dimensional stimuli and neural responses for both efficient coding and Bayesian inference, in this review, we focus on the case of a scalar stimulus value θ and a neural response summarised without loss of information by a scalar internal representation m .

Loss function: a function that specifies the behavioural cost of different estimation errors, that is, deviations in an estimate $\hat{S}$ from the stimulus S . Models of behaviour often aim to minimise the average squared-error loss, $(\hat{S} - S)^2$, or more generally $|\hat{S} - S|^\alpha$ for some $\alpha > 0$.

Maximum a posteriori (MAP) estimate: in the context of Bayesian inference, this involves choosing a stimulus value that has the highest probability under the posterior distribution (i.e., the posterior mode).

Mutual information: a measure of codependence between two random variables. Mutual information quantifies how much uncertainty about one variable is reduced, or how much information is gained, by observing the other variable.

Perceptual bias: a systematic deviation of the perceived stimulus value from its true value, typically quantified as the mean difference. Perceptual biases cannot be observed directly but can be inferred based on responses in discrimination or estimation tasks.

Posterior: the probability distribution $p(S|R)$ expressing an observer's strength of belief that the stimulus takes on a particular value S given the sensory input R , obtained using Bayes' rule.

Prior: the probability distribution $p(S)$ expressing the probability with which an observer expects the stimulus to take on a particular value S before receiving sensory input about it.

Stimulus value: the relevant physical property or feature of a stimulus, on a continuous scale, for example, the angle of an oriented edge relative to horizontal.

Visual working memory: limited capacity cognitive system for temporarily storing and manipulating visual information in support of behaviour.

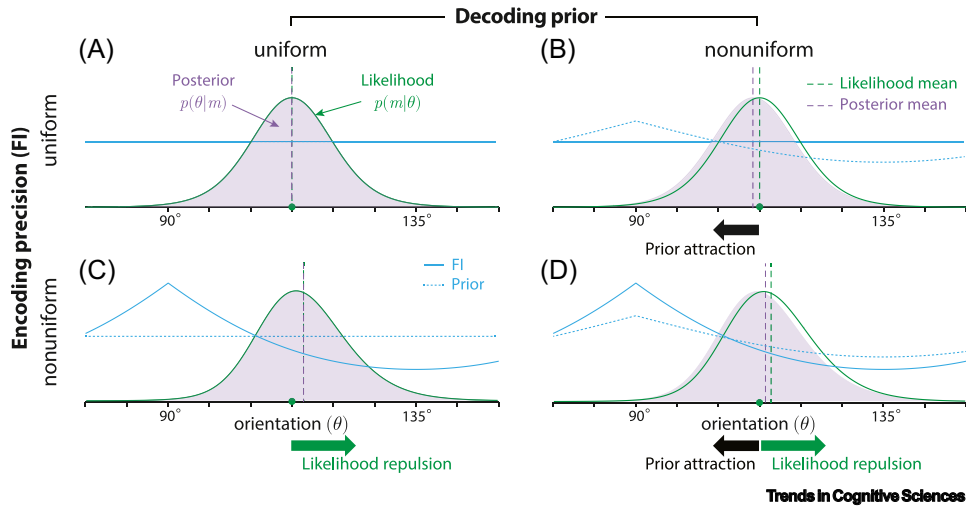


Figure 2. Consequences for estimation of uniform versus nonuniform encoding precision and decoding prior. (A) Uniform encoding and prior yield symmetric likelihoods and posteriors, so the expected estimate matches the stimulus. (B) Uniform encoding with a nonuniform prior illustrates the standard Bayesian account, with estimates attracted towards the prior peak. (C) Nonuniform encoding (as in Figure 1D) produces an asymmetric likelihood with a long tail away from the high-precision region; combined with a uniform prior, the estimate is biased in the same direction. (D) With nonuniform encoding and a nonuniform prior [46], likelihood repulsion and prior attraction combine, producing net estimates that can be biased away from the prior peak. FI: Fisher information.

allocation of resources leads to the following relationship between the prior probability of a stimulus value, $p(\theta)$, and its **discrimination threshold**, $D(\theta)$:

$$D(\theta) \propto \frac{1}{p(\theta)} \quad [1]$$

Thus, stimulus values that are encountered more often can be more finely discriminated from neighbouring features, leading to gradients of **discriminability** across feature space.

According to this framework, the gradients of discriminability caused by efficient coding give rise to systematic **perceptual biases** in Bayesian estimates [45,46]. Consider a stimulus feature in a part of the stimulus space where FI is changing smoothly (green stimulus in Figure 1C,D): it can be discriminated more accurately from stimuli encoded with higher FI (e.g., blue stimulus) than from those with lower FI, in the opposite direction (e.g., red stimulus). This asymmetry is summarised in a **likelihood function** (black curve in Figure 1D, bottom right) that is skewed in the direction of decreasing FI. Figure 2D illustrates the consequences for Bayesian estimation. When the skewed likelihood (green curve) is combined with the nonuniform prior (dotted blue curve) via **Bayes' rule** (Box 2), the **posterior** (purple shaded area) shifts in the direction of increasing prior probability, while inheriting some of the likelihood's asymmetry. For comparison, Figure 2A–C illustrates individual effects of nonuniformity in encoding precision and prior probability on the posterior distribution.

Skewness of the posterior induces a bias in point estimates of the stimulus value, the direction and strength of which depend on the **loss function**. If we define loss $L_\alpha \propto |\text{error}|^\alpha$, then for $\alpha \geq 1$, the overall bias is in the direction opposite to the effect of the prior. For example, under the commonly assumed squared-error loss ($\alpha = 2$), the perceptual estimate is the posterior mean, which is displaced in the direction of decreasing prior probability. This bias can be described

Box 3. Efficient resource allocation

Efficient resource allocation can be formalised using the FI ($J(\theta)$), which quantifies precision in representing a stimulus value θ . Stated simply, FI measures how faithfully a stimulus value is represented in observable sensory responses and, consequently, how precisely it can be decoded from those responses. Higher FI implies greater certainty about the estimated parameter and reduced variability across repeated estimates of θ . The formal relation between FI and the variance of an estimator $\hat{\sigma}^2(\theta)$ is given by the Cramér–Rao bound:

$$\hat{\sigma}^2(\theta) \geq \frac{1}{J(\theta)} \quad [\text{I}]$$

Thus, an unbiased estimator has variance of at least $1/J(\theta)$. For a biased estimator, the Cramér–Rao bound generalises to:

$$\hat{\sigma}^2(\theta) \geq \frac{[1 + b'(\theta)]^2}{J(\theta)} \quad [\text{II}]$$

where $b'(\theta)$ expresses the amount by which the estimation bias changes around θ . Assuming Gaussian variability, it has been shown [49,50] that the discrimination threshold $D(\theta)$ for a stimulus variable can be expressed in terms of estimation bias and variability, and thus related to FI:

$$D(\theta) \propto \frac{\hat{\sigma}(\theta)}{1 + b'(\theta)} \propto \frac{1}{\sqrt{J(\theta)}} \quad [\text{III}]$$

where the second step requires an assumption that the estimation variance satisfies the Cramér–Rao bound (Equation II) or at least exceeds it by a fraction that is constant across stimulus features.

If estimation noise is small, the classic efficient coding goal of maximising mutual information under resource constraints (Box 1) can be achieved by setting FI proportional to the square of the stimulus probability [11,12,48]. Combining this allocation with the result of Equation III, a direct relationship between discrimination thresholds and the observable environmental statistics was derived [46]:

$$J(\theta) \propto p(\theta)^2 \Rightarrow D(\theta) \propto \frac{1}{p(\theta)} \quad [\text{IV}]$$

with the RHS (Equation 1 of the main text) following from Equation III.

Following from this, a relationship between perceptual bias and the shape of the prior distribution was also derived [46,47]. Specifically, the bias at stimulus value θ is proportional to the derivative of the inverse squared prior:

$$b(\theta) \propto \left[\frac{1}{p(\theta)^2} \right]' \quad [\text{V}]$$

where $[\cdot]'$ denotes gradient, that is, how much the quantity in brackets changes with θ . Re-expressing this relationship in terms of discrimination thresholds $D(\theta)$ leads to Equation 2 in the main text.

as anti-Bayesian because it superficially contradicts the classical view that combining likelihood with a prior always pulls estimates towards values with higher prior probability.

Formalising these results, a concise and general perceptual law was derived, relating the perceptual bias with which the stimulus value θ is estimated, $b(\theta)$, and its discriminability [47] (Box 3):

$$b(\theta) \propto [D(\theta)^2]' \quad [2]$$

predicting that perceptual bias varies in direct proportion to the gradient of the squared discrimination threshold. For loss exponents $\alpha \geq 1$, the proportionality constant is positive, so the overall bias is in the ‘anti-Bayesian’ direction of decreasing prior probability. In the limit case of an all-or-nothing loss (L_0), for which the **maximum a posteriori (MAP) estimate** is optimal, Equation 2 also holds but with a negative constant of proportionality, and the direction of bias is reversed. Together, these theoretical results link representational fidelity and perceptual inference in a

precise quantitative framework, offering testable predictions that can be evaluated across a wide range of perceptual tasks.

Evidence from cognitive and perceptual phenomena

The computational principles of the original theory have been framed mostly in terms of low-level visual processing. However, emerging evidence suggests that the same principles can be applied to a diverse range of tasks and behaviours. Here, we review this evidence.

Psychophysical measurement of discriminability and bias

A natural starting point for evaluating predictions of the Bayesian efficient coding framework is to examine domains in which nonuniformity of the sensory prior is already well characterised, such as the case of visual orientation [1,51]. Perceptual accuracy in these domains is assessed using psychophysical methods, including **discrimination** and **estimation tasks** [52–54]. All such tasks require observers to make a comparison: either of a stimulus against an internal standard (e.g., subjective vertical) or, more commonly, of one stimulus with respect to another. As a result, measures of variability or discriminability obtained in such tasks reflect the combined effects of noise in two feature representations. Even more significantly, measures of bias reflect the difference in internal bias between two representations. Comparisons between stimuli perceived with equal bias do not show evidence of bias (failure to take this into account is referred to as the El Greco fallacy [55]).

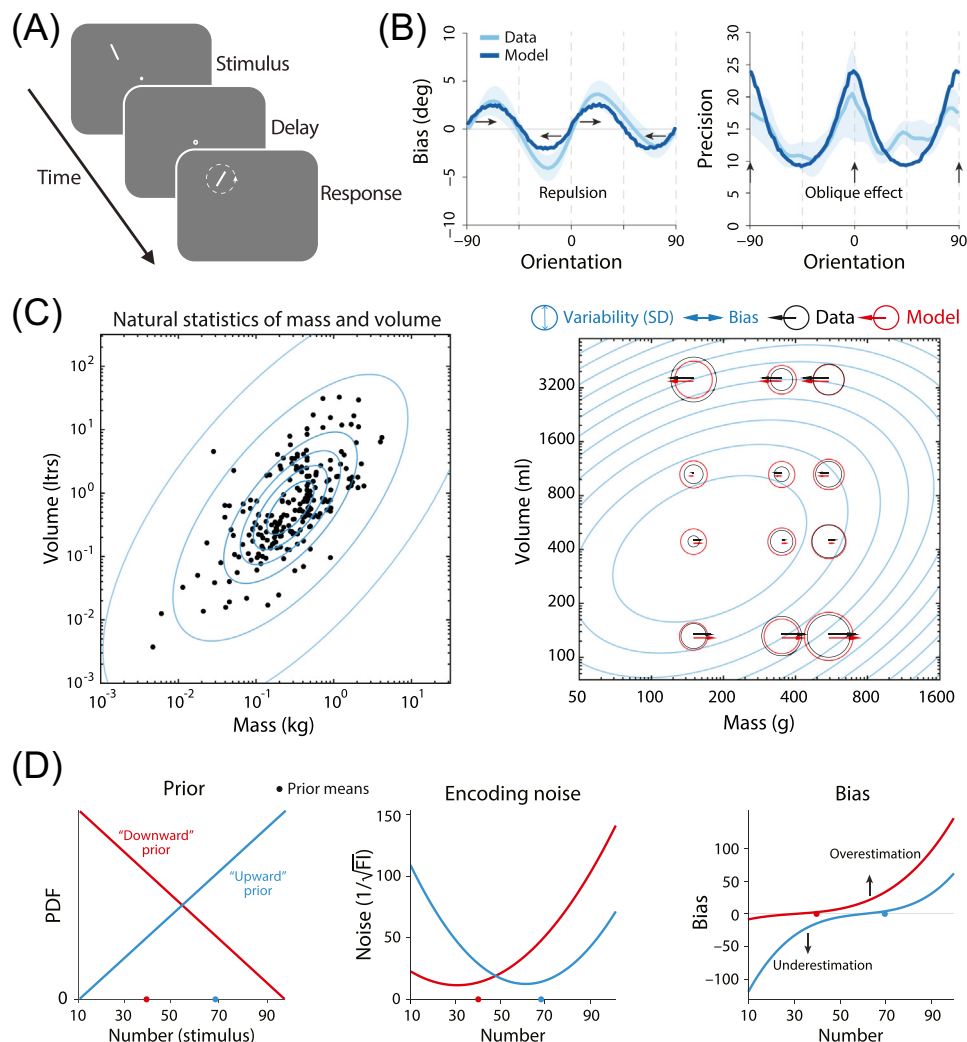
For these reasons, experimental tests of the predictions of Bayesian efficient coding for perception are based on comparisons between stimuli that differ in their expected levels of internal precision. In some cases, this involves an explicit manipulation of noise level, such as when participants report the relative tilt of two simultaneously visible stimuli that differ in experimentally controlled uncertainty [1]. In other cases, the comparison is between two stimuli presented sequentially, separated by a brief interval; an example is the delayed orientation estimation task depicted in Figure 3A. The assumption (sometimes implicit) in such tasks is that the noise associated with representing the first stimulus in **visual working memory** [60–64] is sufficiently large in comparison to the noise in perceiving the visible stimulus, so that the latter can be discounted, and bias and variability in the reproduced feature can be interpreted directly as pertaining to the internal representation of the first stimulus.

Visual perception and memory

Human observers demonstrate superior estimation and discrimination performance for cardinal relative to oblique orientations—a well-established phenomenon known as the ‘oblique effect’ [1,56,65,66]. As qualitatively predicted by the unified framework, behavioural estimates also display systematic repulsive biases away from the cardinal axes—a finding consistently observed in perceptual and delayed estimation tasks [67–73]. Observers also exploit natural statistics when judging orientations of natural images [74] and report the highest confidence for judgements about cardinal orientations [75,76]. Efficient coding principles embedded within a population coding model quantitatively fit these systematic variations in precision and bias in delayed estimation data [56] (Figure 3B).

Motion direction statistics [77,78] and behavioural performance show the same cardinal–oblique anisotropies as orientation, including enhanced performance for cardinal directions [78–81] and systematic repulsive biases away from cardinals [81–83].

Systematic biases also occur in the perception of motion speed, which is typically underestimated, particularly when sensory evidence is weak [84]. Recent work has argued that



Trends in Cognitive Sciences

Figure 3. Evidence for Bayesian efficient coding in perceptual and cognitive tasks. (A) In an orientation delayed estimation task, observers view an oriented bar and after a delay, rotate a second bar to reproduce the remembered orientation. (B) Bias and precision patterns obtained from the task in A. Bias (left) follows a sinusoidal pattern, with reproduced angles deviating away from the nearest cardinal axis. Precision (right) peaks at cardinals, consistent with the oblique effect. Darker lines show the Bayesian efficient coding model fit. Adapted from [56]. (C) Size–weight illusion. Correlation between mass and volume in a sample of liftable everyday objects (left [57]). Right panel shows human observers' SDs (black circle diameters) and mean relative biases (black arrows) demonstrating the size–weight illusion (bias to perceive objects of larger volume as having smaller mass) along with predictions of the fitted multidimensional efficient coding model (red). Contour lines indicate the estimated joint prior over mass and volume. Adapted from [58]. (D) In a magnitude estimation task, observers view two number sequences on each trial and judge which has the larger average. In separate blocks, numbers are drawn from different prior distributions (left)—one favouring lower numbers ('downward') and one favouring higher numbers ('upward'). Fitting a Bayesian efficient coding model to behaviour reveals that numbers are encoded with noise that depends on their prior probability, with less probable numbers represented more noisily (middle). The minima of these functions align with prior means (coloured dots). The model predicts that biases (right) follow the derivative of the inverse Fisher information (i.e., noise variance) (see Equation V in Box 3), producing repulsion from the prior mean. Middle and right panels show predictions of the model that best fitted behaviour. Adapted from [59].

such biases reflect efficient coding of motion and has extended the Bayesian efficient coding model with a prior favouring slow speeds [85], consistent with earlier proposals of slow-speed priors in motion perception [86,87]. This approach not only accounted for behavioural data but also revealed a close correspondence between the fitted priors and those estimated from middle temporal (MT) population responses, linking the model's assumptions to neural implementation (Box 4).

A surprising perceptual effect, which may reflect the same principles, is that under low-pressure sodium light, which removes most colour information [103], faces appear green rather than the reddish tones typically expected for human skin. This paradoxical effect has also been explained by efficient coding constraints on Bayesian perception [104].

Spatial perception and memory also exhibit systematic biases and variability in precision. When estimating heading direction from optic flow, observers show enhanced precision near straight-ahead directions, alongside robust biases across heading space [79,105]. A Bayesian efficient observer model captures these behavioural patterns [106], using a distribution favouring straight-ahead and straight-back directions that aligns with measured heading statistics [107].

Box 4. The neural basis of efficient coding

Although in this review, our focus is on the consequences of efficient coding for perception, and in particular on its measurable effects on behaviour, that is, responses in perceptual decision-making (discrimination or estimation) tasks, in its original form, efficient coding was a theory of neural responses [3]. Indeed, the general principle of efficient coding has successfully accounted for neural response properties in a broad range of sensory systems [4,88–90] and for various forms of learning and short-term adaptation at the neural level [17,91–93].

The specific formulation of the efficient coding principle pursued by Bayesian efficient coding can be most easily translated into neural predictions via a population coding view of sensory processing. Under this view, multiple neurons together code a single (typically scalar) stimulus variable, with each neuron being characterised by its tuning to this variable, summarised by its tuning curve (mean response as a function of the stimulus). The FI provided about the stimulus by such a population can be mathematically derived from the tuning curve properties of the population. In the simplest case (independent Poisson firing of neurons), it is

$$J(\theta) = \sum_i \frac{[f'_i(\theta)]^2}{f_i(\theta)} \quad [1]$$

where i indexes the neurons of the population $f_i(\theta)$ is the tuning curve of neuron i , and $f'_i(\theta) = \frac{d}{d\theta} f_i(\theta)$ is the slope of the tuning curve [94].

As we saw, efficient coding places constraints on $J(\theta)$, the left-hand side of Equation 1, in terms of how it should be related to the stimulus distribution (Equation 1 in the main text, see also [48]). In general, there may be multiple different sets of tuning curves that achieve the same $J(\theta)$ according to Equation 1, and information maximisation without meaningful constraints can easily lead to degenerate predictions about tuning curves [94,95]. For example, increasing the gain of tuning curves, decreasing their width, increasing their density around a given θ , or some combination thereof may all increase $J(\theta)$ and thus provide a neural counterpart to increasing the 'perceptual resources' allocated to representing θ [46,96]. Therefore, identifying which mechanisms are employed is challenging because distinct population codes can produce equivalent representational geometries and discriminability patterns [97]. For example, similar patterns of bias and discriminability in orientation and heading direction can be observed with the highest or the lowest neuronal density at peak discriminability, respectively [98,99]. Unique predictions about neural tuning can only be derived by specifying constraints on some tuning properties [100].

Nevertheless, studies across sensory modalities have revealed that neural tuning properties systematically reflect environmental statistics. In early visual cortex, more neurons prefer cardinal orientations and exhibit narrower tuning at these values [98,101]. Critically, encoding precision (FI) inferred from neural population activity closely matches estimates derived from behaviour [85,102], validating the framework's link between neural codes and perceptual performance.

Similar signatures are observed in spatial memory: locations near scene landmarks are systematically misremembered as being closer to the landmark than they actually were. While traditionally attributed to prior-driven attraction, these biases naturally emerge from the efficient reallocation of encoding resources according to scene structure [108]. Using a Bayesian efficient coding model, which concentrates encoding precision near landmarks, this study also correctly predicted enhanced discrimination at these locations, showing that both bias and precision can be explained by the same encoding mechanism. The study also provides a generalisation of the original theory by applying it to a two-dimensional spatial case.

Multi-feature perception

Most Bayesian efficient coding models have focused on single stimulus variables, such as orientation. However, the occurrences of different features within objects in the natural environment are often correlated, meaning each conveys information about the other. Extending the Bayesian efficient coding framework to paired correlated variables, recent work showed that the empirical relationship between the mass and volume of everyday objects predicts perceptual biases consistent with the size–weight illusion, the well-known anti-Bayesian phenomenon in which smaller objects feel heavier upon lifting than larger objects of the same weight [58] (Figure 3C). The same principles also accounted for the related material–weight illusion. There are many possible implementations of multidimensional Bayesian efficient coding [109], and while several solutions were found to be consistent with behaviour in [58], more theoretical and experimental work is needed to identify which is implemented in the brain (Box 4).

Numerical and value-based cognition

Evidence for Bayesian efficient coding extends to numerical processing. A longstanding puzzle in numerosity research is that small quantities tend to be overestimated, whereas larger ones are underestimated. Although a standard Bayesian model has been proposed as an explanation for this ‘central tendency’ in estimation [110], it does not specify how noise arises in numerosity representations or how such noise shapes estimation biases. Recent work has shown that a capacity-limited system, in which encoding and decoding are adapted to **environmental statistics**, can account for the observed biases, as well as for the reduction in bias as stimulus-induced noise decreases (e.g., with increased exposure time) [111]. A Bayesian efficient coding observer incorporating a log-normal prior suited to natural numerical distributions, and operating under capacity constraints, reproduces these characteristic estimation patterns, demonstrating how biases in number processing can be traced to the same efficient coding principles that govern perception more broadly (see also [112]). Similarly, when comparing averages of numerical streams, humans encode numbers with noise that depends on stimulus probability, leading to systematic biases that are well captured by a Bayesian efficient coding model incorporating the prior distribution of stimuli [59] (Figure 3D).

The representation of value has long been studied within an efficient coding framework [113]. Building on this foundation, a formal model rooted in the Bayesian efficient coding framework has been developed to examine human preference-based decisions [114]. The model used participants’ food ratings to derive individual priors, which informed both the encoding and decoding stages in value-based choice tasks. This model successfully captured systematic patterns of bias, variability, and confidence observed in behaviour. Other studies have also shown that the sensitivity of subjective valuations depends on the distribution of values experienced within a given experimental context [25, 115, 116].

Changes in bias amplitudes with noise

The unified framework predicts that noise introduced at different stages of stimulus processing will produce either repulsive or attractive biases. According to the original formulation of the

framework [46], internal noise corrupts stimulus information once it has been transformed into an efficient internal representation, whereas external noise exerts its effect at an earlier stage, before this transformation. Operationally, external noise is symmetric with respect to the stimulus space, while internal noise is asymmetric, reflecting the nonuniform distribution of neural resources. Thus, increasing internal noise is predicted to accentuate the asymmetry of likelihood functions and thereby strengthen repulsive biases. Increasing external noise broadens the likelihood function without adding to its asymmetry, increasing the influence of the prior at decoding and leading to stronger attractive biases. Despite these clear theoretical predictions, empirical support remains mixed.

In psychophysical tasks, reducing stimulus contrast [87,117,118] or duration [67,119–121] is thought to effectively increase the influence of internal noise by attenuating the signal-to-noise ratio. Consistent with this, several studies have found that repulsive biases increase with shorter exposure times [67,114]. Bias magnitude also scales with the fidelity of cortical decoding, as evidenced by a positive correlation between the strength of repulsive bias and fMRI measures of representational uncertainty [73]. However, conflicting evidence has emerged. In a task using categorical judgements, reducing certainty through manipulations of orientation variance, contrast, or stimulus duration has been shown to reduce or eliminate repulsive biases relative to cardinal orientations [122]. Moreover, manipulations that are thought to increase noise in working memory representations—larger set sizes and extended retention intervals—have not been found to consistently modulate bias amplitude, despite clear increases in overall variability [108,123–125].

Several studies have varied the dispersion of component elements of a stimulus in judgements based on their average [1,126]. This can be interpreted as an external noise manipulation, assuming observers treat the components as random samples and aim to estimate the population mean. While some experiments produced the expected attraction bias with increased dispersion [1], others reported inconsistent results depending on whether dispersion was added to the reference or the probe stimulus [126].

Another potential challenge for the framework is the preliminary, but seemingly consistent, finding [127,128] that tilting the head dissociates orientation biases (repulsive from body- or world-centred cardinals) from variability (most precise for head- or eye-centred cardinals).

Theoretical advances and controversies

The foundational studies of the framework established key principles for linking sensory statistics to perceptual biases. Subsequent work has expanded and refined the framework, including relaxing assumptions about coding objectives to show that lawful bias-discriminability relationships emerge more generally, as well as clarifying the mechanistic origins of attraction versus repulsion to generate more precise predictions.

Generalising coding objectives and constraints

The perceptual law relating bias and discriminability was derived based on the goal of maximising **mutual information** ("infomax") in the low-noise regime under a particular constraint on total FI $\int \sqrt{J(\theta)} d\theta \leq C$ that is invariant under a reparameterisation of the stimulus [46,47]. This constraint limits the total encoding capacity that can be distributed across the stimulus space and implies a necessary trade-off if FI is to be increased for particular stimulus values. Under this constraint, the optimal solution yielded an efficient allocation of FI proportional to the square of the prior (Equation 1). Although infomax is the canonical goal of efficient coding [129], other studies have considered alternative objectives and constraints, showing that different optimisation objectives yield distinct allocations of coding resource with respect to the prior [48].

Specifically, this work derived efficient population codes based on a constraint on total expected activity (population gain) for infomax, ‘discrimax’ (minimising expected squared discrimination threshold, $D(\theta)^2$), and a generalisation to arbitrary powers of $D(\theta)$.

Building on these foundations, it was subsequently shown that minimising an L_α loss function is, under mild assumptions, equivalent to minimising the expected FI raised to a power, $J(\theta)^{-\alpha/2}$, and that under the more general constraint $\int J(\theta)^\beta d\theta \leq C$, this is achieved by setting $J(\theta) \propto p(\theta)^q$ with $q = 1/(\alpha/2 + \beta)$ [130]. Crucially, this work showed that the lawful relationship between perceptual bias and discriminability (Equation 2) is not unique to information-theoretically optimal codes but emerges for FI allocated based on any power q of the prior (‘power-law efficient codes’). Finally, this work also derived an explicit expression for bias under the MAP decoder, and analysed the breakdown of the linear relationship between $b(\theta)$ and $[D(\theta)^2]'$ at higher noise levels (also illustrated in Figure 4).

The power-law efficient codes enable consideration of encoding objectives beyond maximising representational accuracy. One proposal is that encoding might be optimised for a trade-off between accuracy and behavioural cost of errors [66,131]. For example, reducing the exponent q redistributes encoding resources more evenly, shifting some resources away from highly probable stimuli towards less probable ones [66]. The intuition is that, although some stimuli occur rarely, errors for those stimuli may turn out to be particularly costly; consequently, it may be beneficial to sacrifice some accuracy for frequent stimuli to avoid costly errors on infrequent ones. This ‘fitness-maximisation’ solution—closely related to the task-dependency principle of information compression [131]—has received empirical support in psychophysical tasks that directly manipulate the costs or rewards associated with responses [66] (Yang, Q. *et al.*, 2025, *NeurIPS, conference*).

Components of bias

A complementary line of work has focused on refining the efficient coding account of systematic biases. Focusing on the L_2 estimator (the posterior mean), it was shown that the overall bias in this

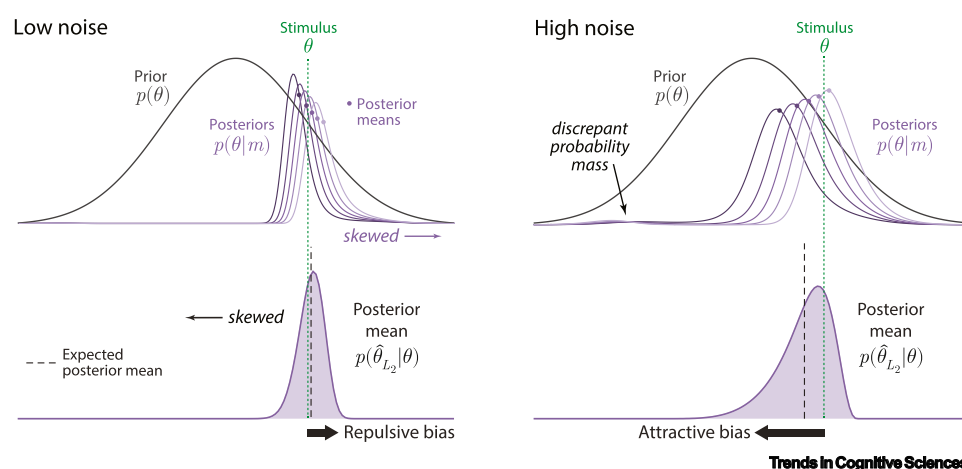


Figure 4. Illustration of the effect of noise level on bias. Under low noise, the prior is approximately linear in the domain of the likelihood, yielding the predicted small repulsive bias (left). As noise increases, the idealised Bayesian efficient coding prediction breaks down (right): the likelihood spreads across the prior peak, making stimulus values on the far side plausible, which shifts the mean posterior back towards the prior peak, producing a strong attractive bias. Simulations based on a Gaussian prior (vertical scale exaggerated) and $J(\theta) \propto p(\theta)^2$. The example illustrates one instance in which the model’s predictions break down at high noise.

estimator can be decomposed into two approximately additive terms arising from the gradients of encoding precision and decoding prior [132]. Likelihood asymmetry shifts estimates in the direction of decreasing encoding precision, $J(\theta)$, while Bayesian integration shifts them in the direction of increasing prior density, $p(\theta)$. For efficient codes, encoding precision increases with increasing prior density, so these components act in opposite directions. For the general class of power-law efficient codes $J(\theta) \propto p(\theta)^q$, the aggregate direction of the bias depends on q . For $q < 1$, the bias in the L_2 estimator is in the direction of increasing prior probability, the usual characteristic of Bayesian integration, while for $q > 1$, the bias is in the opposite anti-Bayesian direction.

This result was subsequently generalised [133] to the wider family of L_α loss functions ($\alpha \geq 1$), giving the approximation

$$b(\theta) = \frac{\alpha + 2}{4} \left[\frac{1}{J(\theta)} \right]' + \frac{1}{J(\theta)} [\log p(\theta)]' \quad [3]$$

where α is the integer loss exponent (Figure 5A). This decomposition shows that bias increases linearly with the exponent α of the loss function. Notably, $p(\theta)$ in this equation refers only to the decoding prior, allowing bias to be predicted in cases where the decoding prior is different from the stimulus distribution that determines the allocation of encoding resources.

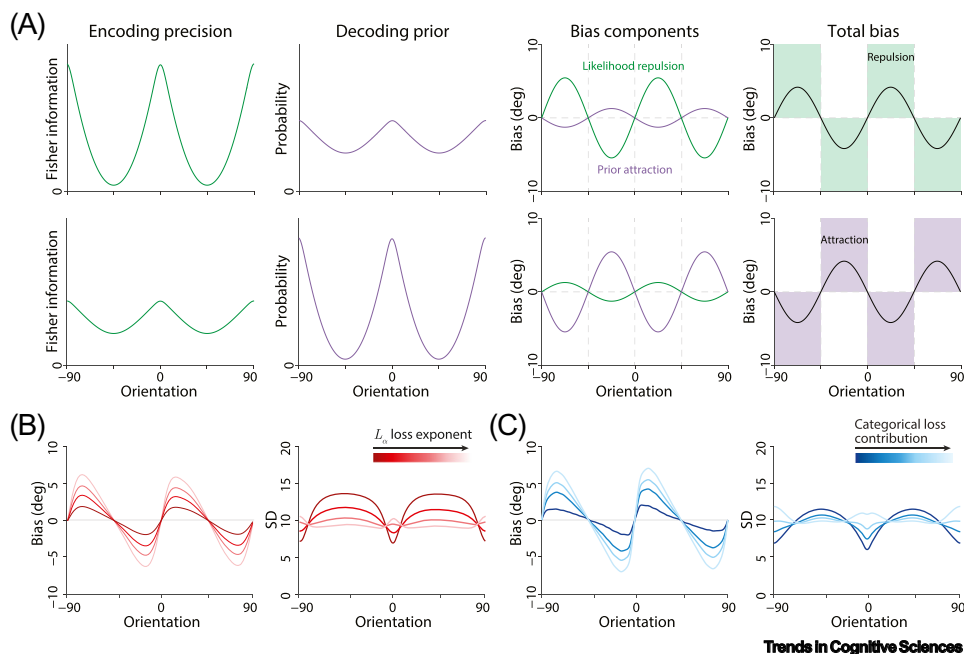


Figure 5. Determinants of bias direction in Bayesian efficient coding. (A) Bias arises from the joint influence of encoding precision and prior density [132,133]. Encoding precision drives repulsive biases away from regions of high FI, whereas the prior drives attraction towards high-probability regions. Net bias can be decomposed into repulsive and attractive terms (Equation 3), which depend on the relative rates of change of these two factors [133]. Rapidly changing FI with slowly varying prior produce net repulsion (top row), whereas slowly varying FI with rapidly changing prior produce net attraction (bottom row). Predicted biases are shown for the L_2 estimator—higher loss exponents amplify repulsion and can attenuate prior attraction even when the prior varies more strongly. (B,C) Predicted bias (left) and standard deviation (right) in an orientation estimation task as a function of L_α loss exponent (B) [133] or categorical loss weight (w_c) (C). Increasing either α or w_c amplifies repulsive biases away from cardinal orientations and inverts the variability pattern, shifting peak variability from oblique to cardinal orientations. FI: Fisher information. Adapted from [133,134].

Explaining empirical bias and variability patterns

Psychophysical studies reveal several observations that are difficult to reconcile with the original theory: perceptual biases are often substantially larger than predicted under the conventional L_2 loss, and patterns of bias and variance show inconsistencies across studies—for instance, both minimal and maximal variance at cardinals have been observed along with repulsive biases in orientation [56,67,70]. One successful approach to fitting larger-than-expected biases involves increasing the loss exponent (e.g., $\alpha \approx 8$), which amplifies likelihood-driven repulsion and thus produces stronger repulsive biases [133] (Figure 5B). Intuitively, higher loss exponents tolerate small errors but strongly penalise larger ones, shifting the optimal estimate towards the posterior tail in order to minimise catastrophic errors in the direction of decreasing prior probability. This also reshapes the variability profile—in orientation estimation tasks, it shifts peak variability from oblique towards cardinal orientations, capturing empirical patterns that differ from the predictions for L_2 .

Beyond freely fitting loss exponents, this approach also developed a method to independently and nonparametrically fit FI and prior distributions, allowing any combination of decoding prior, encoding precision, and loss exponent to be inferred from behavioural data [133]. This flexibility enables the model to reproduce a wide range of stimulus-specific bias and variability patterns across tasks. However, such flexibility also raises concerns: if both encoding and decoding components are freely fitted, the model risks becoming detached from environmental statistics, thereby limiting its grounding in efficient coding theory.

An alternative approach to these same empirical challenges treats perceptual inference as a process operating across both feature and categorical levels (e.g., colour names or compass directions) [134]. This hierarchical structure allows the model to compute separate posteriors and corresponding losses for features and categories, with the total loss given by a weighted combination of feature-level and category-level losses. In this model, larger-than-expected empirical biases are accounted for by increasing the weight assigned to the categorical loss (Figure 5C). Categorical loss is minimised by shifting estimates towards the centre of the most probable category. For orientation, this means oblique angles are reproduced closer to the centres of their respective oblique categories, effectively shifting estimates further from the nearest cardinal category. Additionally, increasing the weight of the categorical loss shifts the variability profile: in orientation estimation tasks, the largest variability moves from oblique to cardinal orientations, capturing the same empirical patterns addressed by higher loss exponents [133].

Note that an essential ingredient in this account is noise in category membership over and above that induced by noise at the feature-level [134]. This category noise also helps explain puzzling results from a task in which a noisy test stimulus (Gabor array) and a ‘noiseless’ adjustable probe (two dots implying an oriented line) were swapped on half of the trials, such that observers sometimes adjusted the noisy array to match the implied line [126]. In the original theory, this should simply reverse the sign of the error, but instead, human performance showed biases away from the cardinals in both conditions [126]. This is explained by the model with category uncertainty because placing the probe closer to the oblique category centres reduces categorical loss, irrespective of which stimulus has greater feature-level noise, and overriding any opposing feature-level bias [134].

Adaptation to the environment

A notable open question in neuroscience concerns how adaptation and learning shape neural codes that guide behaviour. While there are many theoretical approaches to this question [41,135–137], the Bayesian efficient coding account provides a new perspective by jointly

considering adaptation in encoding and inference. In this framework, both encoding precision and decoding prior are assumed to track environmental statistics. In stationary environments—those experienced over evolutionary and developmental timescales—this assumption is plausible and supported by experimental and neural evidence (Box 4) [1,46,47,98,138].

A more controversial question is whether similar forms of adaptation occur over shorter timescales and, if so, whether they primarily involve changes in encoding, decoding, or coordinated updating of both [23,139,140]. In the original framework [47], encoding and decoding were assumed to adapt in a tightly coupled manner, incorporating identical assumptions about environmental structure. In laboratory settings, environmental statistics are typically defined by the distribution of stimuli within the experiment rather than by natural stimulus statistics, allowing direct tests of how perceptual systems recalibrate to recent experience.

Using this approach, studies have provided evidence for adaptive changes in both encoding and decoding. In one study, observers were presented with orientation stimuli drawn from a distribution incongruent with natural statistics (peaked at obliques) [56]. Following exposure to the oblique-dominated environment, repulsive biases away from cardinals decreased. In the Bayesian efficient coding framework, this implies adaptation at the encoding stage: if only the decoding prior had adapted, biases should have increased, as attraction towards obliques would compound the existing repulsion from cardinals. Similarly, prolonged exposure to a uniform distribution of orientations has been shown to gradually flatten the inferred FI profile and reduce repulsive bias, consistent with encoding resources becoming more evenly distributed [70]. Other work has shown that working memory reallocates capacity according to the statistical regularities of remembered items, consistent with efficient coding predictions [66,141].

Evidence from other tasks similarly implicates adaptive changes in encoding mechanisms during adaptation. In a classical tilt adaptation experiment, short-term exposure to an oblique adaptor sharpened sensitivity for orientations similar to the adaptor (and, to a lesser extent, at the orthogonal orientation) while reducing it elsewhere—this pattern was best explained by the reallocation of limited encoding resources (FI) towards these orientations [142]. Adaptive encoding is likewise evident in tilt illusions, where an oriented surround induces repulsive biases and changes in variability for the estimation of central gratings with a similar orientation [102]. Estimates of FI derived from behavioural data and a population-coding analysis of fMRI signals demonstrated increased encoding precision for gratings similar to the surround, consistent with the rapid reallocation of neural resources towards contextually relevant features.

Similar context-dependent efficient coding may play a role in perceptual averaging tasks, where observers compare the mean of an ensemble (e.g., a set of oriented gratings) to a reference [143,144]. In such a task, elements whose features lie closer to the ensemble mean are typically weighted more strongly. Studies have shown that this robust averaging behaviour can arise from efficient coding tuned to the ensemble's overall statistics: stimuli that fall further from the ensemble mean tend to be infrequent, so they are encoded less precisely and thus relatively underweighted in computing the average [145]. When ensemble statistics were fixed relative to a reference stimulus that changed from trial to trial, evidence for efficient coding was observed when ensemble and reference stimuli were presented simultaneously, but this effect was disrupted when stimuli were presented sequentially, suggestive of near-instantaneous reallocation of encoding resources [146].

While the above results all point to short-term and even trial-by-trial adaptation of encoding, a strikingly different conclusion was reached by analysing the effects on motion perception of

bimodal stimulus distributions peaking near cardinals, obliques, or intermediate motion directions [133]. Fitting models in which encoding precision and decoding prior were decoupled, they found that, at the level of the experimental block, resource allocation remained strongest at cardinals regardless of stimulus distribution, whereas decoding priors adapted to the experimentally imposed statistics. For stimulus distributions in which cardinals were more prevalent, the effects of likelihood repulsion and Bayesian attraction were opposed, but when obliques were more common, these effects reinforced each other, increasing the strength of repulsion from the cardinal directions.

A study of serial dependence—a phenomenon in which the estimation of a current stimulus is biased by preceding stimuli at the same location—also found evidence for a dissociation between encoding and decoding stages [147]. Specifically, it was found that perceptual estimates were simultaneously attracted towards recently presented stimuli and repelled from those encountered further in the past. This pattern was explained using a Bayesian efficient coding model, in which decoding priors updated rapidly based on recent history, producing short-term attraction, whereas efficient encoding integrated information over longer timescales, producing long-term repulsion. Other studies of serial biases similarly argue for separate but correlated temporal dynamics of encoding and decoding components [148].

Concluding remarks

Over the past decade, a wide range of sensory and cognitive phenomena that previously lacked normative explanations have been shown to align with the predictions of Bayesian efficient coding. In some cases, this has involved extending the theory beyond stable environmental statistics, accumulated over evolutionary and developmental timescales, to more volatile environments in which relevant regularities cannot be hard-wired into the neural system. An important open question is how the brain balances such long-term regularities with short-term fluctuations in environments and behavioural goals. This question is conceptually related to perceptual and statistical learning, which also focus on how the brain identifies regularities in sensory input and develops internal representations of them [149–151]. Future work could leverage these similarities to advance our understanding of adaptive efficient coding.

Since the introduction of the framework, substantial progress has been made in characterising the component processes that give rise to perceptual biases, particularly by linking these biases to variations in encoding and decoding. Early theoretical work proposed that priors are implicitly embedded within the structure of the encoding population—specifically, in the distribution and shapes of tuning curves across neurons—thereby removing the need for distinct populations to represent sensory evidence and priors [48], and predicting coordinated adaptation of encoding and decoding to environmental statistics [47]. More recent computational work has instead dissociated the contributions of encoding and decoding processes to perceptual biases, demonstrating that likelihood-driven repulsion and prior-driven attraction reflect computationally separable components [132,133]. A major challenge for future work is to determine whether these computationally dissociable components have separable neural substrates, or whether they are jointly implemented as originally proposed [152–155]. One approach to address this question involves developing ‘**image-computable**’ models of sensory areas, which can naturally implement both the encoding and inference components of the framework (Box 5).

Previously underemphasised aspects of the framework—particularly the role of loss functions that formalise behavioural goals and the degree of optimisation of internal representations for a given task—offer a promising route for integrating diverse findings across domains and experimental tasks [133,134]. While the framework has already demonstrated broader applicability

Outstanding questions

How broadly does the unified framework apply across sensory modalities? Despite evidence for efficient coding and Bayesian inference in multiple modalities, systematic tests of their unified operation remain largely confined to vision.

How can we best account for the finding that behavioural biases are larger than predicted and sometimes inconsistent with the original unified theory? Different modifications have been proposed to account for these effects, but at a cost to the parsimony of the original theory.

What mechanisms enable near-instantaneous, efficient adaptation, and how is efficiency balanced between immediate goals and long-term environmental statistics? If neural codes can adapt so rapidly to the current context, why do we see any signature of background statistics in behaviour?

Under what conditions can a system adapt to a particular feature distribution? Which mechanisms mediate such adaptation? Can unsupervised exposure alone drive such adaptation, or is a specific task required for encoding mechanisms to adjust accordingly? How much experience with an environment does an organism need in order to adapt?

Do the encoder and decoder learn the same statistical regularities at the same rate, and does the decoder have full knowledge of adaptation-induced changes in the encoder, such that it interprets the encoded response according to the correct post-adaptation encoding model? The original theory assumes that encoding and decoding rely on identical assumptions during adaptation, but this has not been rigorously tested.

Do likelihood repulsion and prior attraction—key computational components in models of perceptual bias—have separable and identifiable neural signatures?

How can the theory be extended beyond the low-noise regime? Current predictions are derived under low-noise assumptions, yet are applied seemingly successfully to explain

Box 5. Image-computable models: unifying efficient coding and Bayesian inference

While Bayesian efficient coding treats the efficient encoding and Bayesian inference steps as conceptually separate, there is a class of neural models of Bayesian inference that perform both at the same time—at least implicitly. In particular, many ‘image-computable’ models of primary sensory areas formulate neural dynamics as performing inference under a statistically accurate internal model of sensory stimuli [156–158]. Due to their image-computable nature, these models do not only predict how neural responses are tuned to some predefined stimulus feature (e.g., orientation in V1), and how this neural tuning may be adapted to stimulus statistics—as most population coding-based models of efficient coding do. Instead, they provide a more general account for how neural responses depend on arbitrary stimuli, for example, represented as arrays of pixels in the case of vision. This allows them to account for complicated forms of tuning to multiple features that are difficult to incorporate into population coding models (e.g., various extraclassical receptive field effects [157–159]), and to predict why tuning to specific features (e.g., to localised oriented edges) arises in the first place as a result of adapting the brain’s internal model to natural scene statistics [156]. The generative component of image-computable models can be implemented using various architectures, ranging from Gaussian scale mixture models to convolutional neural networks.

Critically, although image-computable models of Bayesian inference are typically not framed in terms of efficient coding, their neural responses can be shown to maximise information with respect to the distribution of stimuli to which their internal model has been adapted [160] (Hinton, G.E. *et al.*, 1993, NeurIPS, conference). Moreover, when their predicted receptive fields are specifically analysed for the distribution of preferred stimulus features (e.g., orientations), they show similar biases (e.g., cardinality biases) as when tuning curves are optimised explicitly for information efficiency in population coding models [161]. (In fact, such effects appear quite generally in image-computable models [162,163], not just in those whose neural responses have been shown to represent Bayesian posteriors.)

In many Bayesian image-computable models, neural responses simply represent the maximum of the Bayesian posterior [156,157], that is, the MAP decoder that is implicitly optimised for a fixed and very specific (and somewhat degenerate) loss function (corresponding to L_{α} in the $\alpha \rightarrow 0$ limit). There also exist generalisations that represent (to some approximation) the full Bayesian posterior [158] (Vértes, E. *et al.*, 2018, NeurIPS, conference), which could, in principle, be combined with any task-dependent loss function by downstream circuits. Beyond classical and extraclassical receptive field properties, these ‘fully Bayesian’ image-computable models also account for a host of other, ubiquitous properties of cortical responses, such as variability, oscillations, stimulus-onset transients, and structured spontaneous activity [164,165]. Whether and how these MAP or fully Bayesian image-computable models would also account for the behavioural signatures of Bayesian efficient coding (e.g., repulsive biases) remain to be tested.

perceptual and memory phenomena where variability is substantial. While attempts have been made to relax the low-noise assumption, a general treatment beyond this regime is lacking.

than initially anticipated, these recent developments also open new theoretical and empirical questions (see [Outstanding questions](#)).

Acknowledgements

This work was supported by the Wellcome Trust (Investigator Award in Science 212262/Z/18/Z to M.L.) and the Croatian Science Foundation (UIP-2025-02-6353 to I.T.).

Declaration of interests

The authors declare no competing interests.

References

- Girshick, A.R. *et al.* (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14, 926–932
- Attneave, F. (1954) Some informational aspects of visual perception. *Psychol. Rev.* 61, 183–193
- Barlow, H. (1961) Possible principles underlying the transformations of sensory messages. In *Sensory Communication* (Rosenblith, W.A., ed.), pp. 216–234. The MIT Press
- Laughlin, S. (1981) A simple coding procedure enhances a neuron’s information capacity. *Z. Naturforsch. C Biosci.* 36, 910–912
- Laughlin, S.B. *et al.* (1998) The metabolic cost of neural information. *Nat. Neurosci.* 1, 36–41
- Attwell, D. and Laughlin, S.B. (2001) An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* 21, 1133–1145
- Hasenstaub, A. *et al.* (2010) Metabolic cost as a unifying principle governing neuronal biophysics. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12329–12334
- Shannon, C.E. (1948) A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423
- MacKay, D.J.C. (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press
- Park, I.M. and Pillow, J.W. (2017) Bayesian efficient coding. *bioRxiv* <https://doi.org/10.1101/178418>
- Brunel, N. and Nadal, J.P. (1998) Mutual information, Fisher information, and population coding. *Neural Comput.* 10, 1731–1757
- Wei, X.-X. and Stocker, A.A. (2016) Mutual information, Fisher information, and efficient coding. *Neural Comput.* 28, 305–326
- van Hateren, J.H. (1992) Real and optimal neural images in early vision. *Nature* 360, 68–70
- Lewicki, M.S. (2002) Efficient coding of natural sounds. *Nat. Neurosci.* 5, 356–363
- Rodríguez, F.A. *et al.* (2010) Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. *J. Neurosci.* 30, 15969–15980

16. Smith, E.C. and Lewicki, M.S. (2006) Efficient auditory coding. *Nature* 439, 978–982
17. Teşileanu, T. *et al.* (2019) Adaptation of olfactory receptor abundances for efficient coding. *Elife* 8, e39279
18. Shao, Y. *et al.* (2020) Compression of dynamic tactile information in the human hand. *Sci. Adv.* 6, eaaz1158
19. Wang, T. *et al.* (2025) Sequential effects in reaching reveal efficient coding in motor planning. *Sci. Adv.* 11, eadv6572
20. Brenner, N. *et al.* (2000) Adaptive rescaling maximizes information transmission. *Neuron* 26, 695–702
21. Dean, I. *et al.* (2005) Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* 8, 1684–1689
22. von der Twer, T. and MacLeod, D.I. (2001) Optimal nonlinear codes for the perception of natural colours. *Network* 12, 395–407
23. Młynarski, W.F. and Hermundstad, A.M. (2018) Adaptive coding for dynamic sensory inference. *Elife* 7, e32055
24. Młynarski, W.F. and Hermundstad, A.M. (2021) Efficient and adaptive sensory codes. *Nat. Neurosci.* 24, 998–1009
25. Summerfield, C. and Tsetsos, K. (2015) Do humans make good decisions? *Trends Cogn. Sci.* 19, 27–34
26. Frydman, C. and Jin, L.J. (2021) Efficient coding and risky choice. *Q. J. Econ.* 137, 161–213
27. Schlütt, H.H. *et al.* (2024) Reward prediction error neurons implement an efficient code for reward. *Nat. Neurosci.* 27, 1333–1339
28. Sims, C.R. (2018) Efficient coding explains the universal law of generalization in human perception. *Science* 360, 652–656
29. Jakob, A.M. and Gershman, S.J. (2023) Rate-distortion theory of neural coding and its implications for working memory. *Elife* 12, e79450
30. Chalk, M. *et al.* (2018) Toward a unified theory of efficient, predictive, and sparse coding. *Proc. Natl. Acad. Sci. U. S. A.* 115, 186–191
31. Manookin, M.B. and Rieke, F. (2023) Two sides of the same coin: efficient and predictive neural coding. *Annu. Rev. Vis. Sci.* 9, 293–311
32. von Helmholtz, H.L.F. (1924) In *Helmholtz's Treatise on Physiological Optics* (James, P.C., ed.), Optical Society of America
33. Knill, D.C., Richards, W., eds (1996) *Perception as Bayesian Inference*, (1st edn) Cambridge University Press
34. Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304
35. Ma, W.J. *et al.* (2023) *Bayesian Models of Perception and Action: An Introduction*, MIT Press
36. Yuille, A. and Kersten, D. (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* 10, 301–308
37. Körding, K.P. and Wolpert, D.M. (2004) The loss function of sensorimotor learning. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9839–9842
38. Jazayeri, M. and Shadlen, M.N. (2010) Temporal context calibrates interval timing. *Nat. Neurosci.* 13, 1020–1026
39. Chalk, M. *et al.* (2010) Rapidly learned stimulus expectations alter perception of motion. *J. Vis.* 10, 2
40. Sotiropoulos, G. *et al.* (2011) Changing expectations about speed alters perceived motion direction. *Curr. Biol.* 21, R883–R884
41. Simoncelli, E.P. and Olshausen, B.A. (2001) Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216
42. Körding, K.P. and Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning. *Nature* 427, 244–247
43. Tassinari, H. *et al.* (2006) Combining priors and noisy visual cues in a rapid pointing task. *J. Neurosci.* 26, 10154–10163
44. Wolpert, D.M. *et al.* (1995) An internal model for sensorimotor integration. *Science* 269, 1880–1882
45. Wei, X.-X. and Stocker, A.A. (2012) Efficient coding provides a direct link between prior and likelihood in perceptual Bayesian inference. In *Advances in Neural Information Processing Systems* (25) (Pereira, F. *et al.*, eds), Curran Associates, Inc
46. Wei, X.-X. and Stocker, A.A. (2015) A Bayesian observer model constrained by efficient coding can explain 'anti-Bayesian' percepts. *Nat. Neurosci.* 18, 1509–1517
47. Wei, X.-X. and Stocker, A.A. (2017) Lawful relation between perceptual bias and discriminability. *Proc. Natl. Acad. Sci. U. S. A.* 114, 10244–10249
48. Ganguli, D. and Simoncelli, E.P. (2014) Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. *Neural Comput.* 26, 2103–2134
49. Seriès, P. *et al.* (2009) Is the homunculus "aware" of sensory adaptation? *Neural Comput.* 21, 3271–3304
50. Seung, H.S. and Sompolinsky, H. (1993) Simple models for reading neuronal population codes. *Proc. Natl. Acad. Sci. U. S. A.* 90, 10749–10753
51. Coppola, D.M. *et al.* (1998) The distribution of oriented contours in the real world. *Proc. Natl. Acad. Sci. U. S. A.* 95, 4002–4006
52. Prinz, W. *et al.* (1998) Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *J. Exp. Psychol.* 24, 261–282
53. Wilken, P. and Ma, W.J. (2004) A detection theory account of change detection. *J. Vis.* 4, 1120–1135
54. Tomić, I. *et al.* (2024) Dissecting the components of error in analogue report tasks. *Behav. Res. Methods* 56, 8196–8213
55. Firestone, C. (2013) On the origin and status of the "El Greco fallacy". *Perception* 42, 672–674
56. Taylor, R. and Bays, P.M. (2018) Efficient coding in visual working memory accounts for stimulus-specific variations in recall. *J. Neurosci.* 38, 7132–7142
57. Peters, M.A.K. *et al.* (2015) Smaller = denser, and the brain knows it: natural statistics of object density shape weight expectations. *PLoS One* 10, e0119794
58. Bays, P.M. (2024) Weight illusions explained by efficient coding based on correlated natural statistics. *Commun. Psychol.* 2, 125
59. Prat-Carrabin, A. and Woodford, M. (2022) Efficient coding of numbers explains decision bias and noise. *Nat. Hum. Behav.* 6, 1142–1152
60. Baddeley, A. (2012) Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29
61. Bays, P.M. *et al.* (2024) Representation and computation in visual working memory. *Nat. Hum. Behav.* 8, 1016–1034
62. D'Esposito, M. and Postle, B.R. (2015) The cognitive neuroscience of working memory. *Annu. Rev. Psychol.* 66, 115–142
63. Ma, W.J. *et al.* (2014) Changing concepts of working memory. *Nat. Neurosci.* 17, 347–356
64. Pasternak, T. and Greenlee, M.W. (2005) Working memory in primate sensory systems. *Nat. Rev. Neurosci.* 6, 97–107
65. Appelle, S. (1972) Perception and discrimination as a function of stimulus orientation: the "oblique effect" in man and animals. *Psychol. Bull.* 78, 266–278
66. Schaffner, J. *et al.* (2023) Sensory perception relies on fitness-maximizing codes. *Nat. Hum. Behav.* 7, 1135–1151
67. de Gardelle, V. *et al.* (2010) An oblique illusion modulated by visibility: non-monotonic sensory integration in orientation processing. *J. Vis.* 10, 6
68. Jastrow, J. (1892) Studies from the university of Wisconsin: on the judgment of angles and positions of lines. *Am. J. Psychol.* 5, 214
69. Lennie, P. (1971) Distortions of perceived orientation. *Nat. New Biol.* 233, 155–156
70. Noel, J.-P. *et al.* (2021) Individuals with autism spectrum disorder have altered visual encoding capacity. *PLoS Biol.* 19, e3001215
71. Pratte, M.S. *et al.* (2017) Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 6–17
72. Shin, H. *et al.* (2017) The effects of delay duration on visual working memory for orientation. *J. Vis.* 17, 10
73. van Bergen, R.S. *et al.* (2015) Sensory uncertainty decoded from visual cortex predicts behavior. *Nat. Neurosci.* 18, 1728–1730
74. A-Izzeddin, E.J. *et al.* (2024) The influence of natural image statistics on upright orientation judgements. *Cognition* 242, 105631
75. Geurts, L.S. *et al.* (2022) Subjective confidence reflects representation of Bayesian probability in cortex. *Nat. Hum. Behav.* 6, 294–305

76. West, R.K. *et al.* (2025) Priors for natural image statistics inform confidence in perceptual decisions. *Conscious. Cogn.* 128, 103818
77. Bex, P.J. *et al.* (2005) Critical band masking in optic flow. *Network* 16, 261–284
78. Dakin, S.C. *et al.* (2005) An oblique effect for local motion: psychophysics and natural movie statistics. *J. Vis.* 5, 878–887
79. Gros, B.L. *et al.* (1998) Anisotropies in visual motion perception: a fresh look. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 15, 2003
80. Chetverikov, A. and Jehee, J.F.M. (2023) Motion direction is represented as a bimodal probability distribution in the human visual cortex. *Nat. Commun.* 14, 7634
81. Wong, W. and Price, N.S.C. (2014) Testing neuronal accounts of anisotropic motion perception with computational modelling. *PLoS One* 9, e113061
82. Rauber, H.-J. and Treue, S. (1998) Reference repulsion when judging the direction of visual motion. *Perception* 27, 393–402
83. Krukowski, A.E. and Stone, L.S. (2005) Expansion of direction space around the cardinal axes revealed by smooth pursuit eye movements. *Neuron* 45, 315–323
84. Stocker, A.A. and Simoncelli, E.P. (2006) Noise characteristics and prior expectations in human visual speed perception. *Nat. Neurosci.* 9, 578–585
85. Zhang, L.-Q. and Stocker, A.A. (2022) Prior expectations in visual speed perception predict encoding characteristics of neurons in area MT. *J. Neurosci.* 42, 2951–2962
86. Durant, S. and Zanker, J.M. (2012) Variation in the local motion statistics of real-life optic flow scenes. *Neural Comput.* 24, 1781–1805
87. Weiss, Y. *et al.* (2002) Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604
88. Atick, J.J. and Redlich, A.N. (1992) What does the retina know about natural scenes? *Neural Comput.* 4, 196–210
89. Bell, A.J. and Sejnowski, T.J. (1997) The “independent components” of natural scenes are edge filters. *Vis. Res.* 37, 3327–3338
90. Doi, E. *et al.* (2012) Efficient coding of spatial information in the primate retina. *J. Neurosci.* 32, 16256–16264
91. Fairhall, A.L. *et al.* (2001) Efficiency and ambiguity in an adaptive neural code. *Nature* 412, 787–792
92. Angeloni, C.F. *et al.* (2023) Dynamics of cortical contrast adaptation predict perception of signals in noise. *Nat. Commun.* 14, 4817
93. Schoups, A. *et al.* (2001) Practising orientation identification improves orientation coding in v1 neurons. *Nature* 412, 549–553
94. Dayan, P. and Abbott, L.F. (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*, Computational Neuroscience, Massachusetts Institute of Technology Press
95. Bethge, M. *et al.* (2002) Optimal short-term population coding: when Fisher information fails. *Neural Comput.* 14, 2317–2351
96. Harrison, W.J. *et al.* (2023) Neural tuning instantiates prior expectations in the human visual system. *Nat. Commun.* 14, 5320
97. Kriegeskorte, N. and Wei, X.-X. (2021) Neural tuning and representational geometry. *Nat. Rev. Neurosci.* 22, 703–718
98. Mansfield, R.J.W. (1974) Neural basis of orientation perception in primate vision. *Science* 186, 1133–1135
99. Gu, Y. *et al.* (2010) Decoding of MSTd population activity accounts for variations in the precision of heading perception. *Neuron* 66, 596–609
100. Ganguli, D. and Simoncelli, E.P. (2016) Neural and perceptual signatures of efficient sensory coding. *arXiv* <https://doi.org/10.48550/arXiv.1603.00058>
101. Li, B. *et al.* (2003) Oblique effect: a neural basis in the visual cortex. *J. Neurophysiol.* 90, 204–217
102. Zhang, L.-Q. *et al.* (2025) The tilt illusion arises from an efficient reallocation of neural coding resources at the contextual boundary. *Proc. Natl. Acad. Sci. U. S. A.* 122, e2421565122
103. Boynton, R.M. and Purl, K.F. (1989) Categorical colour perception under low-pressure sodium lighting with small amounts of added incandescent illumination. *Light. Res. Technol.* 21, 23–27
104. Hasantash, M. *et al.* (2019) Paradoxical impact of memory on color appearance of faces. *Nat. Commun.* 10, 3010
105. Cuturi, L.F. and MacNeillage, P.R. (2013) Systematic biases in human heading estimation. *PLoS One* 8, e56862
106. Sun, Q. *et al.* (2025) A linear perception-action mapping accounts for response range-dependent biases in heading estimation from optic flow. *PLoS Comput. Biol.* 21, e1013147
107. Sinnott, C. and MacNeillage, P. (2019) Statistical characterization of heading stimuli in natural environments using SLAM. *J. Vis.* 19, 53
108. Langlois, T.A. *et al.* (2021) Serial reproduction reveals the geometry of visuospatial representations. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2012938118
109. Yerxa, T.E. *et al.* (2020) Efficient sensory coding of multidimensional stimuli. *PLoS Comput. Biol.* 16, e1008146
110. Anobile, G. *et al.* (2012) Linear mapping of numbers onto space requires attention. *Cognition* 122, 454–459
111. Heng, J.A. *et al.* (2025) Efficient numerosity estimation under limited time. *PLoS Comput. Biol.* 21, e1012790
112. Heng, J.A. *et al.* (2020) Efficient sampling and noisy decisions. *Elife* 9, e54962
113. Louie, K. and Glimcher, P.W. (2012) Efficient coding and the neural representation of value. *Ann. N. Y. Acad. Sci.* 1251, 13–32
114. Polania, R. *et al.* (2019) Efficient coding of subjective value. *Nat. Neurosci.* 22, 134–142
115. Payzan-lenestour, E. and Woodford, M. (2018) “Outlier blindness”: efficient coding generates an inability to represent extreme values. *SSRN Electron. J.* Published online March 29, 2018. <https://doi.org/10.2139/ssrn.3152166>
116. Woodford, M. (2020) Modeling imprecision in perception, valuation, and choice. *Annu. Rev. Eco.* 12, 579–601
117. Bays, P.M. (2016) A signature of neural coding at human perceptual limits. *J. Vis.* 16, 4
118. Tomić, I. and Bays, P.M. (2018) Internal but not external noise frees working memory resources. *PLoS Comput. Biol.* 14, e1006488
119. Bays, P.M. *et al.* (2011) Temporal dynamics of encoding, storage, and reallocation of visual working memory. *J. Vis.* 11, 6
120. Shibuya, H. and Bundesen, C. (1988) Visual selection from multielement displays: measuring and modeling effects of exposure duration. *J. Exp. Psychol.* 14, 591–600
121. Tomić, I. and Bays, P.M. (2024) A dynamic neural resource model bridges sensory and working memory. *Elife* 12, RP91034
122. Patten, M.L. *et al.* (2017) Correlates of perceptual orientation biases in human primary visual cortex. *J. Neurosci.* 37, 4744–4750
123. Bae, G.-Y. (2021) Neural evidence for categorical biases in location and orientation representations in a working memory task. *Neuroimage* 240, 118366
124. Rademaker, R.L. *et al.* (2018) Evidence of gradual loss of precision for simple features and complex objects in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 925–940
125. Tomić, I. *et al.* (2025) Dynamics of variability and bias in working memory. *PsyArXiv* https://doi.org/10.31234/osf.io/y2bjv_v1
126. Tomassini, A. *et al.* (2010) Orientation uncertainty reduces perceived obliquity. *Vis. Res.* 50, 541–547
127. Rademaker, R. *et al.* (2017) Dissociable biases in orientation recall: the oblique effect follows retinal coordinates, while repulsion from cardinal follows real-world coordinates. *J. Vis.* 17, 107
128. Servetnik, M.V. *et al.* (2025) Visual representations in the human brain rely on a reference frame that is in between allocentric and retinocentric coordinates. *bioRxiv* <https://doi.org/10.1101/2025.10.25.684527>
129. Wang, Z. *et al.* (2016) Efficient neural codes that minimize L_p reconstruction error. *Neural Comput.* 28, 2656–2686
130. Morais, M. and Pillow, J.W. (2018) Power-law efficient neural codes provide general link between perceptual bias and discriminability. In *Advances in Neural Information Processing Systems* (31), Curran Associates, Inc
131. Bates, C.J. and Jacobs, R.A. (2020) Efficient data compression in perception and perceptual memory. *Psychol. Rev.* 127, 891–917
132. Prat-Carrabin, A. and Woodford, M. (2021) Bias and variance of the Bayesian-mean decoder. *arXiv* <https://doi.org/10.48550/ARXIV.2105.13705>

133. Hahn, M. and Wei, X.-X. (2024) A unifying theory explains seemingly contradictory biases in perceptual estimation. *Nat. Neurosci.* 27, 793–804
134. Mao, J. and Stocker, A.A. (2024) Sensory perception is a holistic inference process. *Psychol. Rev.* 131, 858–890
135. Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192
136. Wainwright, M.J. (1999) Visual adaptation as optimal information transmission. *Vis. Res.* 39, 3960–3974
137. Weber, A.I. *et al.* (2019) Coding principles in adaptation. *Annu. Rev. Vis. Sci.* 5, 427–449
138. Furmanski, C.S. and Engel, S.A. (2000) An oblique effect in human primary visual cortex. *Nat. Neurosci.* 3, 535–536
139. Roach, N.W. *et al.* (2017) Generalization of prior information for rapid Bayesian time estimation. *Proc. Natl. Acad. Sci. U. S. A.* 114, 412–417
140. Berniker, M. *et al.* (2010) Learning priors for Bayesian computations in the nervous system. *PLoS One* 5, e12686
141. Bates, C.J. *et al.* (2019) Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *J. Vis.* 19, 11
142. Mao, J. *et al.* (2025) Adaptation optimizes sensory encoding for future stimuli. *PLoS Comput. Biol.* 21, e1012746
143. De Gardelle, V. and Summerfield, C. (2011) Robust averaging during perceptual judgment. *Proc. Natl. Acad. Sci. U. S. A.* 108, 13341–13346
144. Vickie, L. *et al.* (2017) Robust averaging protects decisions from noise in neural computations. *PLoS Comput. Biol.* 13, e1005723
145. Ni, L. and Stocker, A.A. (2023) Efficient sensory encoding predicts robust averaging. *Cognition* 232, 105334
146. Ni, L. and Stocker, A.A. (2026) Dynamics of efficient ensemble coding. *Cognition* 271, 106432
147. Fritsche, M. *et al.* (2020) A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. *Elife* 9, e55389
148. Van Geert, E. *et al.* (2024) An efficient Bayesian observer model of attractive and repulsive temporal context effects when perceiving multistable dot lattices. *J. Vis.* 24, 18
149. Fiser, J. *et al.* (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130
150. Fiser, J. and Lengyel, G. (2022) Statistical learning in vision. *Annu. Rev. Vis. Sci.* 8, 265–290
151. Szpiro, S.F.A. *et al.* (2025) Perceptual learning improves discrimination but does not reduce distortions in appearance. *PLoS Comput. Biol.* 21, e1012980
152. Prat-Carrabin, A. *et al.* (2025) Distributed range adaptation in human parietal encoding of numbers. *bioRxiv* <https://doi.org/10.1101/2025.09.25.675916>
153. Sheehan, T.C. and Serences, J.T. (2022) Attractive serial dependence overcomes repulsive neuronal adaptation. *PLoS Biol.* 20, e3001711
154. Malerba, S.B. *et al.* (2024) Jointly efficient encoding and decoding in neural populations. *PLoS Comput. Biol.* 20, e1012240
155. Sohn, H. and Narain, D. (2021) Neural implementations of Bayesian inference. *Curr. Opin. Neurobiol.* 70, 121–129
156. Olshausen, B.A. and Field, D.J. (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609
157. Rao, R.P.N. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87
158. Orbán, G. *et al.* (2016) Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron* 92, 530–543
159. Schwartz, O. and Simoncelli, E.P. (2001) Natural signal statistics and sensory gain control. *Nat. Neurosci.* 4, 819–825
160. Cardoso, J.-F. (2002) Infomax and maximum likelihood for blind source separation. *IEEE Signal Process. Lett.* 4, 112–114
161. Berkes, P. *et al.* (2009) A structured model of video reproduces primary visual cortical organisation. *PLoS Comput. Biol.* 5, e1000495
162. Benjamin, A.S. *et al.* (2022) Efficient neural codes naturally emerge through gradient descent learning. *Nat. Commun.* 13, 7972
163. Henderson, M. and Serences, J.T. (2021) Biased orientation representations can be explained by experience with nonuniform training set statistics. *J. Vis.* 21, 10
164. Echeveste, R. *et al.* (2020) Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. *Nat. Neurosci.* 23, 1138–1149
165. Berkes, P. *et al.* (2011) Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* 331, 83–87