

Research



Cite this article: Marzen SE, DeDeo S. 2017 The evolution of lossy compression. *J. R. Soc. Interface* **14**: 20170166. <http://dx.doi.org/10.1098/rsif.2017.0166>

Received: 5 March 2017
Accepted: 18 April 2017

Subject Category:
Life Sciences—Physics interface

Subject Areas:
biomathematics, computational biology, biocomplexity

Keywords:
lossy compression, rate–distortion, information theory, perception, signalling, neuroscience

Author for correspondence:
Simon DeDeo
e-mail: sdedeo@andrew.cmu.edu

Sarah E. Marzen^{1,2} and Simon DeDeo^{3,4}

¹Physics of Living Systems, Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
²Redwood Center for Theoretical Neuroscience, and Department of Physics, University of California at Berkeley, Berkeley, CA 94720, USA
³Department of Social and Decision Sciences, Carnegie Mellon University, 5000 Forbes Avenue, BP 208, Pittsburgh, PA 15213, USA
⁴Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

SD, 0000-0002-5346-9393

In complex environments, there are costs to both ignorance and perception. An organism needs to track fitness-relevant information about its world, but the more information it tracks, the more resources it must devote to perception. As a first step towards a general understanding of this trade-off, we use a tool from information theory, rate–distortion theory, to study large, unstructured environments with fixed, randomly drawn penalties for stimuli confusion (‘distortions’). We identify two distinct regimes for organisms in these environments: a high-fidelity regime where perceptual costs grow linearly with environmental complexity, and a low-fidelity regime where perceptual costs are, remarkably, independent of the number of environmental states. This suggests that in environments of rapidly increasing complexity, well-adapted organisms will find themselves able to make, just barely, the most subtle distinctions in their environment.

1. Introduction

To survive, organisms must extract useful information from their environment. This is true over an individual’s lifetime, when neural spikes [1], signalling molecules [2,3] or epigenetic markers [4] encode transient features, as well as at the population level and over generational timescales, where the genome can be understood as hard-wiring facts about the environments under which it evolved [5]. Processing infrastructure may be built dynamically in response to environmental complexity [6–9], but organisms cannot track all potentially useful information because the real world is too complicated and information bottlenecks within an organism’s sensory system prevent the transmission of every detail. In the face of such constraints, they can reduce resource demands by tracking a smaller number of features [10–14].

Researchers have focused on how biological sensory systems might optimize functionality given resource constraints, or, conversely, minimize the resources required to accomplish a task. In neuroscience, the efficient coding hypothesis [15] and minimal cortical wiring hypothesis [16] are two well-known and highly influential examples. Much of this work has focused on maximizing the ratio of transmitted information (usually, the mutual information between stimulus and spike train) to the energy required for such a code, e.g. [17–24].

Researchers often talk about the creation of more efficient codes that find clever ways to pass information from sensors to later stages in the brain’s processing system by adapting to changes in environmental statistics [25–27] and selectively encoding meaningful features of the environment [28]. However, there is much less understanding of what happens when an organism responds to environmental complexity by discarding information, rather than finding ways to encode it more efficiently.

In this paper, we use an information-theoretic tool—rate–distortion theory—to quantify the trade-off between acquisition costs and perceptual distortion [29]. Rate–distortion allows us to talk about the extent to which an organism can save costs of transmitting a representation of their environment by selectively discarding information.

When they do such a compression, evolved organisms are expected to structure their perceptual systems to avoid dangerous confusions (not mistaking tigers for bushes) while strategically containing processing costs by allowing for ambiguity (using a single representation for both tigers and lions)—a form of *lossy* compression that avoids transmitting unnecessary and less-useful information. Our use of this paradigm connects directly to recent experimental work in the cognitive sciences that uses rate–distortion theory to study errors made in laboratory perception exercises [30].

Lions and bushes are informal examples from mammalian perception, but similar concerns apply to, for example, a cellular signalling system which might need to distinguish temperature signals from signs of low pH, while tolerating confusion of high temperature with low oxygenation. We include both low-level percepts and the higher-level concepts they create and that play a role in decision-making [31–33]. Transmission costs include error-correction and circuit redundancy necessary in noisy systems [34], as well as costs associated with more basic trade-offs associated with volume and energy resources and that appear even when internal noise is absent [16,35–37].

Our work applies both to the problem of how an organism gathers information from the environment (the problem of sensory ecology), and how that information is transferred within the organism’s cognitive apparatus from one unit to another (as happens, for example, in communication between internal encoder and decoder units in what [38] describes as ‘Marr’s motif’ [39]).

In §2, we briefly review rate–distortion theory and how it provides a model of the trade-offs between perception and accuracy for biological organisms. In §3, we show that this general formalism predicts the existence of two distinct regimes in which an organism can operate. In what we refer to as the ‘high-fidelity’ regime, where an organism places a premium on accurate representations, the resources an organism requires to achieve that accuracy grow in tandem with environmental complexity, potentially without bound. This is true even when the organism accepts a certain level of error. However, if the organism is able to tolerate a critical level of error in its representations, this picture changes drastically. In this ‘low-fidelity’ regime, the resources required to achieve a given perceptual accuracy are fixed even when the environment grows in complexity. In §4, we discuss the application of these results to the understanding of perceptual systems in the wild.

2. Theoretical framework

There are both fitness costs and fitness benefits to accurate sensory perception ([40] and references therein). It is costly to set-up the neural architectures that promote increased perceptual ability, for example, and it is also costly to use them. On the other hand, increased perceptual ability can increase fitness, by (for example) allowing the organism to avoid predators, or more efficiently gather scarce resources.

Well-adapted organisms will necessarily balance these costs and benefits, but it is not obvious how this is to be quantified. An organism that has more efficient coding mechanisms will be able to perceive more information in the environment than one that does not.

We tackle this problem in a simplified set-up previously studied in [41]. As in [42], model organisms have (i) an adaptive representation system, which we refer to as a ‘sensory codebook’, and that specifies how an environmental state affects, probabilistically, the internal state of the organism and (ii) two constants: a *rate*, r , at which the perceptual apparatus gains information from the environment, and a *gain*, β , which quantifies the costs of doing so.

While an organism can save energy and effort by devoting only limited resources to perception, it incurs costs from doing so. This is quantified using the *distortion measure* $d(x, \hat{x})$, which specifies the cost incurred by an organism confusing an environmental state x for another state, \hat{x} .

Distortion and rate together dictate an organism’s fitness. We can think (for example) of the average distortion as correlated with the energy that the organism failed to take from the environment, while rate correlates with the energy that the organism expended in making the measurements necessary to make that penalty low. Seen in this way, the organism has two competing goals, wanting to minimize both the costs of gathering information (by throwing out some of it), and the costs of acting on the information that, because it is incomplete, may occasionally be misleading.

If we consider a sensory apparatus of the brain, and fix a particular coding system, then rate r serves as a proxy for neuron number, so that the energy expenditure of this organism’s brain is $\beta^{-1}r$, where β^{-1} is the average rate of energy use for a single neuron [43].

In our simple model here, this enables us to make the assumption that the total cost to the organism is given by this critical quantity, $D + \beta^{-1}r$, and that the fitness of an organism, in an evolutionary sense, is a monotonically decreasing function f of the organism’s total cost. (In general, our results are robust to the further generalization of the fitness cost to a sum of monotonically increasing functions of both D and r ; for simplicity in this paper, we consider the linear case.)

At each generation, the organism chooses its rate and gain prior to any environmental exposure [44]. The environment size N is chosen separately, and a new distortion measure d is drawn from an ensemble of $N \times N$ matrices with probability $p(d)$. We use $\rho(d)$ to denote the probability density function from which entries are drawn at the risk of confusion. An organism with parameters r and β has fitness $f(D + \beta^{-1}r)$ in that generation.

So far, we have defined (i) the costs of perception, through a rate r at which the organism gains information from the environment, (ii) the costs of acting in the presence of incomplete information, specified by the distortion matrix d , and (iii) a parameter β which specifies the particular way an organism balances the two costs against each other: an organism with a larger β finds it easier to process information.

Now, we show how both r and d can be related to a perceptual process. It is simpler to define d first. We say that the organism’s sensory codebook maps each one of the N environmental states, x , to one of N sensory percepts. Further downstream in the organism’s processing of this stimuli, it must then decode this sensory percept into an estimate of

the true state of the environment. When the estimated environmental states are not equivalent to the true input, the organism incurs an average per-symbol cost that we assume to be given by $(1/n) \sum_{i=1}^N d(x_i, \tilde{x}_i)$, where $d(x, \tilde{x})$ is the distortion measure defined above, n is the number of observations, and x_i the i th observation. We assume that there is no cost to perfectly representing an environmental state, $d(x, x) = 0$, i.e. that the distortion measure is normal. Over time, the average cost per input symbol tends to an average distortion D ,

$$D = \sum_{x, \tilde{x}} p(x, \tilde{x}) d(x, \tilde{x}),$$

where $p(\tilde{x}|x)$ describes the probability that an environmental state x (drawn from the set X) is represented by the percept symbol \tilde{x} (which, for simplicity, we can also assume to be also drawn from the set X).

Having defined d , we then define r as the mutual information, $I[X; \tilde{X}]$ between the environment and the perceptual state. Mutual information captures the r of information transmission, by quantifying the drop in uncertainty about the environment from before the percept arrives to after it has been received

$$I[X; \tilde{X}] = H[X] - H[X | \tilde{X}], \quad (2.1)$$

where $H[X]$ is the entropy (uncertainty) of the environment, and $H[X | \tilde{X}]$ is the conditional entropy, given knowledge of the perceptual state \tilde{x} .

When perception is perfect, each x is mapped to a unique \tilde{x} and the second term is zero: on receiving the downstream perceptual data, the organism has zero uncertainty. This means that the information about the environment has been completely transmitted to the organism. However, in the real world, constraints on the rate of information transmission will mean that only some of the information is passed along, and this means that only an approximate mapping is possible. In this case, the information transmitted is lower than what is in the environment, the second term in equation (2.1) will be greater than zero, and, as a consequence the associated codebooks will be noisy. They will occasionally map an environmental state x to the wrong percept (\tilde{x}' , say, instead of \tilde{x}), and D will be non-zero.

We have shown how to relate both r and d to the perceptual process, via a codebook p . Some codebooks, of course, are better than others, and rate–distortion focuses on those codebooks that do best given a particular hard constraint on D (or, equivalently, a hard constraint on r). The rate–distortion function $R(D)$ defines an asymptotically achievable lower bound on the rate required to achieve a given average level of distortion

$$R(D) = \min_{p(\tilde{x}|x): d \leq D} I[X; \tilde{X}]. \quad (2.2)$$

One can interchangeably speak of the distortion–rate function, the inverse of the $R(D)$ function

$$D(R) = \min_{p(\tilde{x}|x): I[X; \tilde{X}] \leq R} \mathbb{E}[d]. \quad (2.3)$$

The $p(\tilde{x}|x)$ which achieve the minima provide a guide as to the kinds of confusions one would expect to see in optimal or near-optimal sensory codebooks [45].

There are intuitive reasons that one might see the rate as a natural resource cost. The most obvious, perhaps, is that, once one fixes an encoding system, the required number of sensory

relay neurons required is also lower-bounded by the rate–distortion function. This bound is weakened if the neurons are not encoding sensory information combinatorially, as may happen in a compressed-sensing situation further downstream of the perceptual apparatus [38]. In the extreme case, where every distinct state is given a separate neuron then $N_{\text{eff}} = 2^{R(D)}$, the (effective) number of states stored by an organism’s codebook at a single time step, might be the more relevant processing cost. Increased neuron number leads to an increase in both preset and operational costs. In either case, there is evidence that an increase in neuron number yields a corresponding increase in the accuracy of the organism’s internal representation of the environment [44]. Alternatively, the rate of heat dissipation needed to store and erase sensory information is at least $k_B T \log(2) I[X; \tilde{X}] \geq k_B T \log(2) R(D)$ [46,47], where T is the ambient temperature and k_B is Boltzmann’s constant. This can be viewed as a (possibly weak) lower bound on the operational cost of memory.

Rate–distortion theory is a departure from other normative efforts in that while the mutual information between stimulus and response quantifies the rate, representational accuracy is quantified by a distortion measure. Transmission rate can either be directly connected to energy or material usage, or understood more generally as a costly thing to evolve [42]. Conversely, the cost of inaccuracy has been neglected. In the past, investigators have assumed that the organism intrinsically cared about information flow [48–50], perhaps implicitly appealing to its operational definitions via the rate–distortion theorem [29], and either showed that empirical codebooks were similar to information-theoretic optimal codebooks [48,49] or inferred an effective distortion measure [30,50].

In some cases [49,51,52], the distortion measure was chosen to be a predictive informational distortion, in which sensory percepts are penalized by their inability to predict future environmental states. Here, we have introduced non-intrinsic environmental structure via a distortion measure $d(x, \tilde{x})$ that dictates the nature of the representations that the system uses.

To calculate the rate–distortion function $R(D)$, we find the codebook $p_\beta(\tilde{x}|x)$ which minimizes the objective function

$$\mathcal{L}_\beta = \beta \sum_{x, \tilde{x}} p(x) p(\tilde{x}|x) d(x, \tilde{x}) + I[X; \tilde{X}], \quad (2.4)$$

by iterating the nonlinear equation

$$p(\tilde{x}|x) = \frac{p(\tilde{x}) e^{-\beta d(x, \tilde{x})}}{Z_\beta(x)}, \quad (2.5)$$

where $Z_\beta(x) = \sum_{\tilde{x}} p(\tilde{x}) e^{-\beta d(x, \tilde{x})}$ as described in [53] for 500 iterations at each β , or 1000 iterations at each β if needed. Because fitness is convex in the choice of the codebook, finding the optimal point corresponds to an unsupervised classification of sensory signals based on gradient descent. It is thus a natural task for an organism’s neural system [54–57], and also corresponds to the simplest models of selection in evolutionary time [58]. From this optimal codebook, we find the corresponding expected distortion D_β and rate R_β as

$$R_\beta = \sum_{x, \tilde{x}} p(x) p(\tilde{x}|x) \log \frac{p(\tilde{x}|x)}{p(\tilde{x})}, = -\beta D_\beta - \sum_x p(x) \log Z_\beta(x) \quad (2.6)$$

and

$$D_\beta = \sum_{x, \tilde{x}} p(x)p(\tilde{x}|x)d(x, \tilde{x}), \quad (2.7)$$

respectively. Expected distortion is at most $D_{\max} = \lim_{\beta \rightarrow 0} D_\beta$. By varying β from 0 to ∞ , we parametrically trace out the rate–distortion function $R(D)$. In our simulations, we choose β 's such that the corresponding D_β 's evenly tile the interval between 0 and $D_{10^{10}} \approx D_{\max}$. Code to perform these calculations, including the construction of the codebook that minimizes the Lagrangian in equation (2.4) given a distortion matrix, is available in [59].

In this article, we assume that the distribution over N possible inputs $p(x)$ is uniform, so $p(x) = 1/N$ for all x . We view $H[X] = \log N$ as a proxy for environmental complexity. In §3, we focus on the case where off-diagonal entries of $d(x, \tilde{x})$ are drawn independently from some distribution $\rho(d)$ with support on $[d_{\min}, \infty)$, and set $d(x, x) = 0$.

Every environment has a ‘minimal confound’, defined as the minimal distortion that can arise from a miscoding of the true environmental input; mathematically, it is defined as

$$d_{\min} := \min_{x \neq \tilde{x}} d(x, \tilde{x}). \quad (2.8)$$

In this paper, we study exponential and lognormal distributions, shifted so that d_{\min} is equal to 0, 1 or 20. These provide minimal models of unstructured environments with many degrees of freedom in which some mistakes are more costly than others. The relationship between an organism's toleration of (average) distortion, and the minimal confound level, will turn out to be the dividing line between the low- and high-fidelity regimes.

3. Results

Evolutionary processes select organisms for higher fitness; the fitness of an organism can be thought of as its reproductive rate relative to average. A higher than average fitness will, given sufficiently faithful reproduction, translate into exponential gains for that organism and its descendants. Rather than considering absolute differences we can then focus on rank order—what changes cause the fitness to rise or fall relative to baseline.

Recall that in our simple set-up, an organism's fitness is some monotonically decreasing function f of the organism's total energetic cost, $D(r) + \beta^{-1}r$, which has three parameters: β is the inverse of the average rate of energy use of a single neuron, referred to as ‘gain’; r is the organism's genetically determined maximal allowable rate; and $D(r)$ is the environment-dependent distortion–rate function at rate r . Here, we have assumed that the sensory codebook has minimized distortion for its given rate, and so set distortion to be the distortion–rate function evaluated at r .

Even simple organisms, like single-cell bacteria, exist in environments in which N is very large, e.g. as in [48]. Practically speaking, then, we are often in the regime in which our resources are dwarfed by the environmental complexity. Naively, we might conclude that in this limit, internal representations of the environment must be quite inaccurate, i.e. $\mathbb{E}[d]$ must become increasingly large. As it turns out, this intuition is half-right. Our results suggest that an organism will only operate and continue to operate at high levels of perceptual accuracy when its gain β grows in tandem with environmental complexity.

The minimal confound d_{\min} defines the boundary between two distinct regimes. In the low-fidelity regime, when $D > d_{\min}$, as environmental complexity grows larger and larger, the resources required to achieve a given distortion D asymptote to a finite constant. In the high-fidelity regime, when $D \leq d_{\min}$, the resources required to achieve that error rate grow without bound.

These results are quite general, but we will begin with concrete examples. We first consider the scenario in which $N^2 - N$ distortions are drawn randomly from some distribution $\rho(d)$. We consider the scaling of the expected rate–distortion function over the distribution of random distortion matrices with the environmental complexity, $\log N$. As argued previously [41], in the large N limit, the rate–distortion function of any given environment becomes arbitrarily close to this expectation value with high probability. This previously described result is apparent in simulation results in figure 1, but so is the above-mentioned presence of two regimes: when $\rho(d)$ is a mean-shifted exponential, rate $R(D)$ appears to grow with environmental complexity $\log N$ only for distortions D below the minimal confound $d_{\min} = 1$.

Simulation results for the mean-shifted exponential and lognormal distributions $\rho(d)$ shown in figures 2 and 3 reveal how quickly rate increases with environmental complexity. In the low-fidelity regime, the rate averaged over the distortion matrix distribution, $\mathbb{E}[R(D)]$, asymptotes to a finite constant as environmental complexity $\log N$ tends to infinity, as shown in figure 2. In the high-fidelity regime, rate $R(D)$ scales linearly with environmental complexity $\log N$, as shown in figure 3. The rate at which expected resources scales with environmental complexity increases as the desired distortion D decreases, appearing to slow to 0 as D approaches d_{\min} .

In the low-fidelity regime, the suggestive results of the simulations can be confirmed by a simple analytic argument. For each of the N environmental states, and expected distortion D , there will be, on average, $NP(d < D)$ states that are allowable ambiguities—i.e. the organism can take as synonyms for that percept. While this is not the most efficient coding, it provides an upper bound to the optimal rate of

$$R(D) \leq \log_2 \frac{1}{P(d < D)}. \quad (3.1)$$

For the exponential, this upper bound means that ambiguities accumulate sufficiently fast that the organism will need at most 6.6 bits (for D equal to 1%) or 10 bits (for D equal to 0.1%) even when the set of environmental states becomes arbitrarily large.

That the bound relies on the conditional distribution function means that our results are robust to heavy-tailed distributions. What matters is the existence of harmless synonyms; for the states we end up being forbidden to confuse, penalties can be arbitrarily large. We can see this in figure 2, where we consider a lognormal distribution with mean and variance chosen so that $P(d < 0.1\%)$ is identical to the exponential case. The existence of large penalties in the lognormal case does not affect the asymptotic behaviour.

These simulation results show not only the validity of the bound, but that actual behaviour has a rapid asymptote; in both the exponential and lognormal cases, the scaling of $R(D)$ is slower than logarithmic in N .

This bound is not tight; as can be seen in figure 2, far better compressions are possible, and the system asymptotes

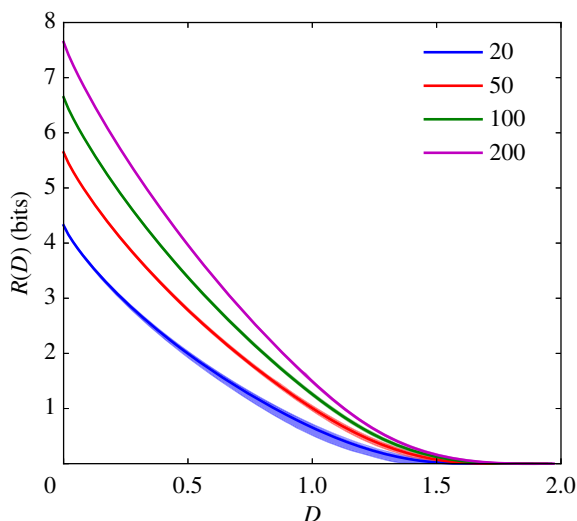


Figure 1. Rate–distortion functions for mean-shifted exponential $p(d)$ for increasing N . Rate–distortion functions for N of 20 (bottom curve), 50, 100 and 200 (top curve), and $p(d) = \begin{cases} 0 & d < 1 \\ e^{-(d-1)} & d \geq 1 \end{cases}$, so that $d_{\min} = 1$. Shaded regions indicate 68% confidence intervals calculated by bootstrapping from 50 samples, and lines within those shaded regions are the means. At $N \geq 100$, rate–distortion functions for different draws from the ensemble are nearly identical. As N grows, $R(D)$ asymptotes to a constant when $D > 1$, but grows without bound when $D < 1$; visually, this means that, reading vertically the $R(D)$ curves for increasing N converge to a limit when D is greater than one, but remain separated when D is below that limit. (Online version in colour.)

at much lower values. Even so, this bound in equation (3.1) implies that the limit of the expected rate–distortion functions in the low-fidelity regime as environmental complexity increases exists.

More generally, in structured environments, $R(D)$ is bounded from above so that $R(D) \leq (1/N) \sum_x \log_2(N/N_D(x))$, where $N_D(x)$ is the number of synonyms for x with distortion penalty less than D . As long as N_D is asymptotically proportional to N , this upper bound implies asymptotic insensitivity to the number of states of the environment.

This remarkable result has, at first, a counterintuitive feel. As environmental richness rises, it seems that the organism should have to track increasing numbers of states. However, the compression efficiency depends on the difference between environmental uncertainty before and after the signal is received, and these two differences, in the asymptotic limit, scale identically with N .

The existence of asymptotic bounds on memory can also be understood by an example from software engineering. As the web grows in size, a search tool such as Google needs to track an increasing number of pages (environmental complexity rises). If the growth of the web is uniform, and the tool is well built, however, the number of keywords a user needs to put into the search query does not change over time. For any particular query—‘information theory neuroscience’, say—the results returned will vary, as more and more relevant pages are created (ambiguity rises), but the user will be similarly satisfied with the results (error costs remain low).¹

As we approach d_{\min} , the upper bound given by $\log_2(1/P(d < D))$ becomes increasingly weak, diverging when $D = d_{\min}$. As we approach this threshold, we enter the

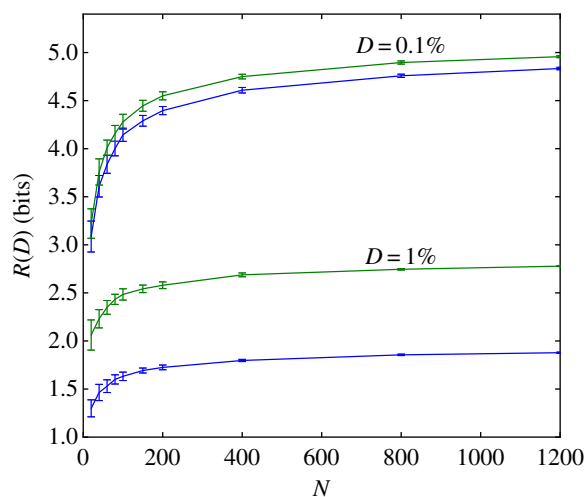


Figure 2. Fault-tolerant organisms can radically simplify their environment. When an organism can tolerate errors in perception, large savings in storage are possible. Shown here is the scaling between environmental size, N , and perceptual costs, $R(D)$, when we constrain the average distortion, D to be 1% and 0.1% of the average cost of random guessing. We consider both environments with exponential and with lognormal (heavy-tailed) fitness, indicated by blue (lower of pair) and green (upper of pair), respectively; in both cases, d_{\min} is equal to zero. A simple argument (see the text) shows that in this low-fidelity regime, and as N goes to infinity, perceptual demands asymptote to a constant. (Online version in colour.)

high-fidelity regime of $D < d_{\min}$, where organisms attempt to distinguish difference at a fine-grained level. In this regime, rate increases without bound as environmental complexity increases.

The apparent linear scaling of $R(D)$ with $\log N$ in the high-fidelity regime shown in figure 3 is confirmed by another simple analytic argument. As in the previous case, we can upper-bound coding costs by construction of a sub-optimal codebook; we can also lower-bound coding costs by finding the optimal codebook for a strictly less-stringent environment. Here, the sub-optimal codebook allocates a total probability C equally to all off-diagonal elements. The rate of this codebook is $\log_2 N - H_b(C) - C \log_2(N - 1)$, while its expected distortion is $C \bar{d}$, where

$$\bar{d} := \sum_x \sum_{\tilde{x} \neq x} \frac{d(x, \tilde{x})}{N(N-1)} \quad (3.2)$$

is the mean off-diagonal distortion. This yields an upper bound on the true rate–distortion function

$$R(D) \leq \log_2 N - \frac{D}{\bar{d}} \log_2(N-1) - H_b\left(\frac{D}{\bar{d}}\right). \quad (3.3)$$

A less-stringent environment is one in which all distortions are d_{\min} , and as $d(x, \tilde{x}) \geq d_{\min}$, this environment has strictly lower resource costs for the same distortion. The rate–distortion function of this environment was given in [60]

$$\log_2 N - \frac{D}{d_{\min}} \log_2(N-1) - H_b\left(\frac{D}{d_{\min}}\right) \leq R(D). \quad (3.4)$$

Together, these place upper and lower bounds on the true rate–distortion function in the high-fidelity regime. In the large N limit, these bounds simplify to

$$\left(1 - \frac{D}{d_{\min}}\right) \log_2 N \leq R(D) \leq \left(1 - \frac{D}{\langle \bar{d} \rangle}\right) \log_2 N, \quad (3.5)$$

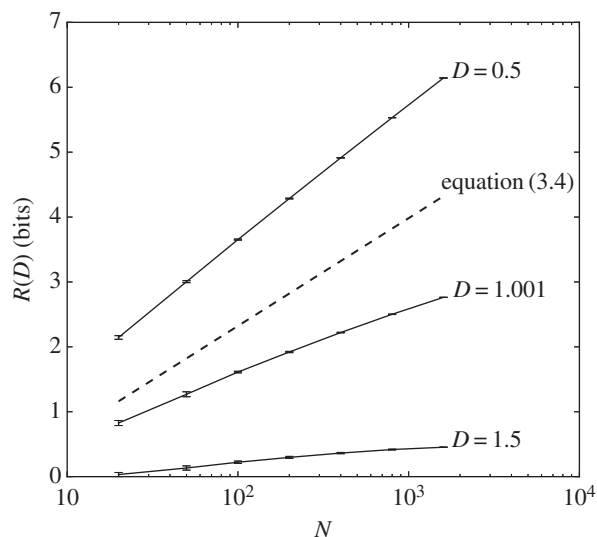


Figure 3. The high-fidelity regime is sensitive to environmental complexity. Shown here are encoding costs for the shifted exponential distribution, where all errors have a minimal penalty, d_{\min} , of one. When we allow average distortions to be greater than d_{\min} , we are in the low-fidelity regime, and costs, as measured by $R(D)$, asymptote to a constant. In the high-fidelity regime, when D is less than d_{\min} , costs continue to increase, such that $R(D)$ becomes linear in $\log N$; the dashed line shows a comparison to the lower bound given in equation (3.4). The change-over in scaling happens at $D = d_{\min}$. Standard errors and means are calculated from 25 samples.

plus small $O(1)$ corrections. In short, the rate scales as $O(\log N)$ in the high-fidelity regime, $D < d_{\min}$, as expected.

These scaling results hold even when off-diagonal entries are partially correlated, as long as correlation scales grow more slowly than environment size so that $N_D \propto N$. As a step towards introducing more structure into these environments, we consider drawing distortion matrices as follows. First, an initial $N \times N$ distortion matrix is constructed by drawing $d(x, \tilde{x})$ i.i.d. from a mean-shifted exponential, $\rho(d) = \begin{cases} 0 & d < 1 \\ e^{-(d-1)} & d \geq 1 \end{cases}$. Next, each entry $d(x, \tilde{x})$ is replaced by $(d(x, \tilde{x}) + d(x, \tilde{x}'))/2$, where $\tilde{x}' \neq \tilde{x}$ is randomly chosen. This final distortion matrix has pairwise-correlated entries. Aforementioned scaling results hold, both using the analytic arguments above and simulation results in figure 4.

Finally, the gain β will at least need to scale as $\log N$ in order to retain D below d_{\min} , as shown in figure 5. A non-coding codebook will have an expected distortion of roughly $\langle d \rangle$ and a rate of 0; a codebook in the high-fidelity regime will have some expected distortion $D < d_{\min}$ and a rate of $C \log_2 N$ for some constant C bounded by $1 - D/d_{\min} \leq C \leq 1 - D/\langle d \rangle$. The high-fidelity codebook outperforms the non-coding codebook when $\beta \langle d \rangle \geq \beta D + C \log_2 N$ or when $\beta \geq (C/(\langle d \rangle - D)) \log_2 N$.

On evolutionary timescales, we expect metabolic efficiency to increase. This is seen in the long-term evolution experiments on *Escherichia coli* as it adapts to new environments [61,62], more generally, it is seen in the macroevolutionary record as the metabolic rate per gram decreases with body size while (in general) body size tends to increase [63].

In the framework described here, the effect of increasing metabolic efficiency is to increase β and thereby to allow the organism to achieve smaller D . When N is large and rising on timescales faster than those on which evolution increases β , we expect organisms to stall out at $D = d_{\min}$.

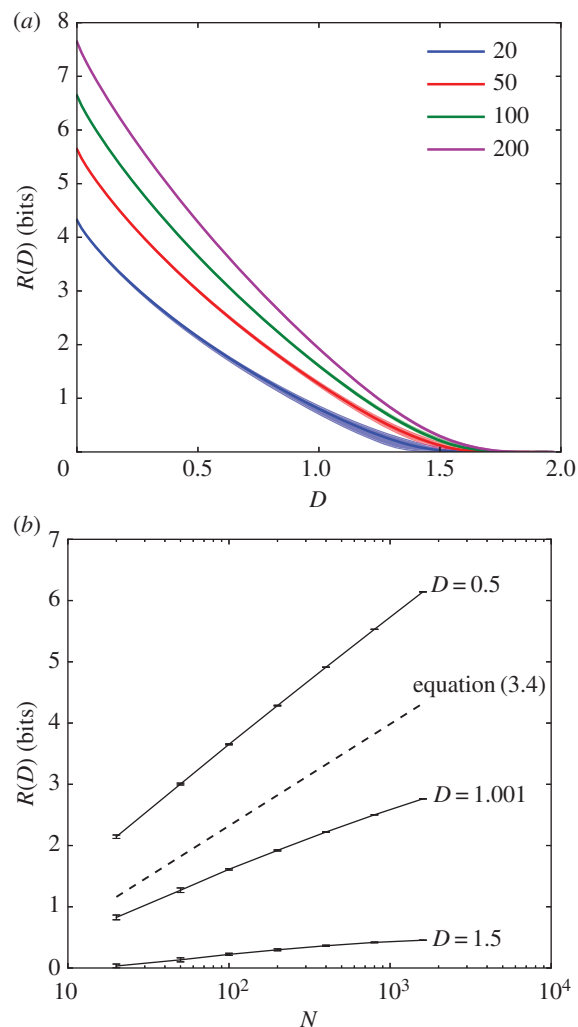


Figure 4. Scaling results for a more structured environment. The distortion matrices are drawn randomly as described in the main text. (a) The rate–distortion function with shaded 68% confidence intervals and mean $\mathbb{E}[R(D)]$ calculated from bootstrapping the rate–distortion functions of 25 different such environments. (b) The scaling of $R(D)$ at various D and N for environments, with means and standard errors calculated from 25 samples. (Online version in colour.)

This is because whenever they go below this error rate, increases in environmental complexity N erase the gains. Even when they are poised around d_{\min} , increasing β will still give an individual an evolutionary advantage; the effect of a changing environment is to force organisms to compete on efficiency, rather than richness of internal representations. These results imply that when organisms compete in an environment of rising complexity, we expect to find their perceptual apparatus poised around $D \sim d_{\min}$. We refer to this as poised perception. Figure 5 shows visually how this driving force to d_{\min} occurs when β rises more slowly than $\log(N)$.

It is useful to relate this poised perception result to the underlying rate–distortion paradigm. Rate–distortion itself can be understood as saying ‘given a particular error tolerance we want to meet, how low can the transmission rate be’. In that sense, the theory can be understood as talking about achieving a ‘just adequate’ level of processing given constraints.

Our finding on poised perception adds a second level of ‘just adequacy’. It says that an organism’s error tolerance

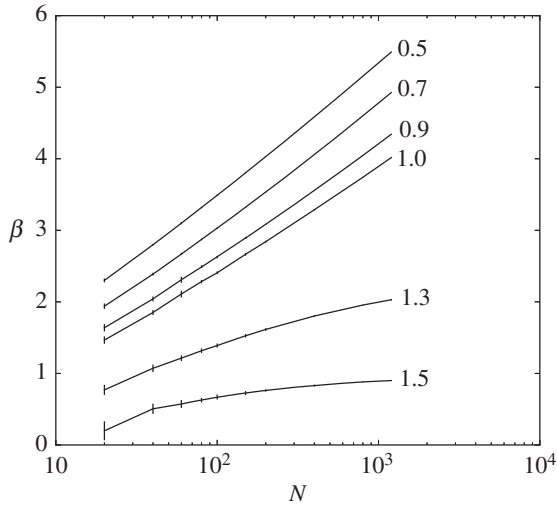


Figure 5. The metabolic efficiency required to achieve a particular distortion in the high-fidelity regime increases with environmental complexity. Shown are plots of the metabolic efficiency, β , required to achieve a particular distortion as a function of environmental complexity, $\log N$, in random environments in which $d(x, \tilde{x})$ is drawn i.i.d. from $p(d) = \begin{cases} 0 & d < 1 \\ e^{-(d-1)} & d \geq 1 \end{cases}$. In the high-fidelity regime, $D < 1$, the required β increases linearly with $\log N$; in the low-fidelity regime, the required β asymptotes to a finite constant. As earlier, means and standard errors were calculated from 25 samples.

will be driven, by evolutionary forces, to a point just adequate for distinguishing minimal confounds. At that point, the organism then devotes perceptual resources ‘just adequate’ to achieving that ‘just adequate’ level of error.

4. Discussion

There are costs and benefits to improved sensory perception. Better tracking of the environment help organisms to acquire new resources. However, the increases in perceptual accuracy needed to achieve this tracking usually require increases in the required size or energy consumption of sensory apparatus. Our work here establishes lower bounds on these trade-offs in terms of the rate–distortion function, which describe the minimal rate required to achieve a given minimal level of perceptual inaccuracy.

We argued that these functions showed surprising regularities in large environments, even though the optimal rate-limited sensory codebooks were highly environment-dependent. Marzen & DeDeo [41] focused on the insensitivity of the rate–distortion function to particular choices of the distortion measure; here, we focus on the scaling of the rate–distortion function with environmental complexity.

Every ensemble of environments has a ‘minimal confound’ d_{\min} , the minimal price that one pays for confusing one environmental state with another. In a low-fidelity regime, when an organism’s average distortion is larger than d_{\min} , increasing environmental complexity does not increase perceptual load. As the number of environmental states increases, innocuous synonyms accumulate. They do so sufficiently fast that an organism can continue to represent the fitness-relevant features within constant memory.

It is only in a high-fidelity regime, when an organism attempts to achieve average distortions below d_{\min} , that

perceptual load becomes sensitive to complexity. High-fidelity representations of the world do not scale, and an organism that attempts to break this threshold will find that, when the number of environmental states increases, it will be driven back to the low-fidelity regime, unless its perceptual apparatus rapidly increases in size.

Together, these results suggest that well-adapted organisms will evolve to a point of ‘poised perception’, where they can barely distinguish objects that are maximally similar.

It is worth considering how this analysis generalizes to the case of a continuum of environmental states. In this case, we have some (usually continuous function) $d(x, \tilde{x})$. To see how this alters the analysis, let us take (as a simple choice) d to be mean squared error, $(x - \tilde{x})^2$. If we assume, again for specificity, that the signal itself is Gaussian with variance σ^2 , then an analytic calculation shows that $R(D)$ is equal to $1/2 \log \sigma^2/D$ when D is less than σ^2 (and zero otherwise). In this case, as long as we track some information about the environment, we are (as expected) never in the ‘low-fidelity’ regime; an increase in environmental complexity (here corresponding to an increase in the range of values the environment can take on, i.e. an increase in σ^2) always leads to an increase in $R(D)$ for fixed D . It is only upon introducing discrete structure in the distortion matrix (e.g. the need to determine both stimulus amplitude, and stimulus type) that the separation between high- and low-fidelity regimes returns.

We have focused on the effect of increasing perceptual abilities due to changes in metabolic efficiency over evolutionary time. However, it is well-known that individuals can increase their perceptual abilities through training in ontogenetic time, as is seen, for example, in video-game players [64–67]. We expect, therefore, the emergence of poised perception on much more rapid timescales. Sims [30] has already demonstrated the utility of rate–distortion theory for the study of perceptual performance under laboratory conditions, and this suggests the possibility of testing the emergence of this phenomenon under a combination of both training (to increase β) and stimulus complexity N .

In other words, when we attempt to make distinctions within a sub-category of increasing size, we will be barely competent at making the least important distinctions. Informally, we can just about distinguish our least-important friends when our circles expand. Such ideas might be tested in a laboratory-based study that trains subjects on a set of increasingly difficult distinction tasks with varying rewards. The results here predict that, as they improve their decision-making abilities, individuals will develop representations that are barely able to make distinctions between the least-important cases. It is also possible to consider the simulation of artificial environments, in an evolutionary robotics or artificial life paradigm [68], to determine the parameter ranges (selection pressure, growth rate of complexity) under which these bounds are achieved.

In many cases, the threats and benefits posed by an environment have a hierarchical structure: there are many threats that are roughly equally bad, and there are many forms of prey, say, that are roughly equally good. Such a hierarchical structure means that we do not expect organisms to have a single system for representing their environment. The results here then apply to sub-domains of the perceptual problem. In the case of predator–prey, for example, we expect optimal codebooks will look block-diagonal, with a clear

predator–prey distinction, and then sub-distinctions for the two categories (eagle versus leopard; inedible versus edible plants).

Similar implications apply in the cognitive and social sciences, where the semantics may play a role in organizing a perceptual hierarchy. A small number of coarse-grained distinctions are often found in human social cognition, where we expect a block-diagonal structure over a small, fixed number of categories (kin versus non-kin, in-group versus out-group [69]). Within each of the large distinctions, our representational systems are then tasked with making fine-grained distinctions over sets of varying size.

For the large distinction, N is fixed, but environmental complexity can increase within each category. As new forms of predators or prey arise, the arguments here suggest that organisms will be poised at the thresholds within each subsystem. When, for example, prey may be toxic, predators will evolve to distinguish toxic from non-toxic prey, but tend not to distinguish between near-synonyms *within* either the toxic or non-toxic categories.

One example of the role played by poised perception in the wild is the evolution of Batesian mimicry—where a non-toxic prey species imitates a toxic one—when such mimicry is driven by the perceptual abilities of predators [70]. Our arguments here suggest that predator perception evolves in such a way that Batesian mimics will have a phenotypic range similar to that found between two toxic species. A harmless Batesian mimic can then emerge if it can approximate, in appearance, a toxic species by roughly the same amount as that species resembles a second species, also toxic. This also suggests that a diversity of equally toxic species will lead to less-accurate mimicry. This is not because a mimic will attempt to model different species simultaneously (the multimodal hypothesis [71]), rather that predators who attempt to make finer-grained distinctions within the toxic-species space will find themselves driven back to d_{\min} when the total number of species increases.

This work is only a first step towards a better understanding of trade-offs between accuracy and perceptual cost in biological systems. Most notably, by reference to a look-up table, or codebook, we neglect the costs of processing the information. The challenges of post-processing, or storing, the output of a perceptual module may argue for adding additional constraints to equation (2.4); see, e.g. [72], where constraints on memory of the input suggest the use of

deterministic codebooks, and the more general discussion of the varieties of computation costs in [14].

5. Conclusion

In any environment, there will be some levels of perceptual accuracy that are nearly impossible for any physical being to achieve. Outside of those effectively forbidden regions, the required size of the sensory apparatus may well depend only on very coarse environmental statistics [41]. Well-adapted organisms are then likely to take advantage of such regularities, allocating resources to gain the information they need to survive.

When they do this, they will encounter not only the constraints expected from the material conditions they experience in both ontogenetic and evolutionary time, but also constraints that arise from the fact that they are, in part, simply information processors in a stochastic but regular world. Our findings suggest that the challenges of perception dramatically increase when organisms attempt to enter a high-fidelity regime. Organisms that are ‘too good’ at perception will find that their abilities rapidly erode when environmental complexity increases, driving them to transition point back into a lower-fidelity regime.

Data accessibility. The code necessary to reproduce the figures in this paper has been released as the Python package, PyRated, and is available at <http://bit.ly/pyrated>.

Authors’ contributions. S.E.M. and S.D. jointly conceived the study, conducted mathematical analysis and wrote the manuscript. S.E.M. wrote the code PyRated used in the paper’s numerical investigations.

Competing interests. We declare we have no competing interests.

Funding. This research was supported by National Science Foundation grant no. EF-1137929. S.E.M. acknowledges the support of a University of California Berkeley Chancellor’s Fellowship, and of an NSF Graduate Research Fellowship.

Acknowledgments. We thank Chris Hillar, John M. Beggs, Jonathon D. Crystal, Michael Lachmann, Jim Crutchfield, Paul Reichers, Susanne Still, Mike DeWeese, Van Savage and anonymous referees for helpful comments.

Endnote

¹We are grateful to Michael Lachmann for suggesting this example.

References

- Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W. 1999 *Spikes: exploring the neural code*. Cambridge, MA: MIT press.
- Tlusty T. 2008 Rate–distortion scenario for the emergence and evolution of noisy molecular codes. *Phys. Rev. Lett.* **100**, 048101. (doi:10.1103/PhysRevLett.100.048101)
- Hancock J. 2010 *Cell signalling*. Oxford, UK: Oxford University Press.
- Prohaska SJ, Stadler PF, Krakauer DC. 2010 Innovation in gene regulation: the case of chromatin computation. *J. Theor. Biol.* **265**, 27–44. (doi:10.1016/j.jtbi.2010.03.011)
- Krakauer DC. 2011 Darwinian demons, evolutionary complexity, and information maximization. *Chaos Interdiscip. J. Nonlinear Sci.* **21**, 037110. (doi:10.1063/1.3643064)
- Beaulieu C, Colonnier M. 1989 Number and size of neurons and synapses in the motor cortex of cats raised in different environmental complexities. *J. Comp. Neurol.* **289**, 178–187. (doi:10.1002/cne.902890115)
- Leggio MG, Mandolesi L, Federico F, Spirito F, Ricci B, Gelfo F, Petrosini L. 2005 Environmental enrichment promotes improved spatial abilities and enhanced dendritic growth in the rat. *Behav. Brain Res.* **163**, 78–90. (doi:10.1016/j.bbr.2005.04.009)
- Simpson J, Kelly JP. 2011 The impact of environmental enrichment in laboratory rats. behavioural and neurochemical aspects. *Behav. Brain Res.* **222**, 246–264. (doi:10.1016/j.bbr.2011.04.002)
- Scholz J, Allemang-Grand R, Dazai J, Lerch JP. 2015 Environmental enrichment is associated with rapid volumetric brain changes in adult mice. *NeuroImage* **109**, 190–198. (doi:10.1016/j.neuroimage.2015.01.027)
- Crutchfield JP. 1994 The calculi of emergence: computation, dynamics, and induction. *Phys.*

- D* **75**, 11–54. (doi:10.1016/0167-2789(94)90273-9)
11. Bialek W, Nemenman I, Tishby N. 2001 Predictability, complexity, and learning. *Neural Comput.* **13**, 2409–2463. (doi:10.1162/089976601753195969)
 12. Sims CA. 2003 Implications of rational inattention. *J. Monetary Econ.* **50**, 665–690. (doi:10.1016/S0304-3932(03)00029-1)
 13. Gershman SJ, Niv Y. 2013 Perceptual estimation obeys Occam's Razor. *Front. Psych.* **4**, 1–11. (doi:10.3389/fpsyg.2013.00623)
 14. Wolpert D, Libby E, Grochow JA, DeDeo S. 2004 Optimal high-level descriptions of dynamical systems. (<https://arxiv.org/abs/1409.7403>)
 15. Barlow HB. 1961 Possible principles underlying the transformations of sensory messages. In *Sensory communication* (ed. WA Rosenblith), pp. 217–234. Cambridge, MA: MIT press.
 16. Chklovskii DB, Koulakov AA. 2004 Maps in the brain: what can we learn from them? *Annu. Rev. Neurosci.* **27**, 369–392. (doi:10.1146/annurev.neuro.27.070203.144226)
 17. Levy WB, Baxter RA. 1996 Energy efficient neural codes. *Neural Comput.* **8**, 531–543. (doi:10.1162/neco.1996.8.3.531)
 18. Olshausen BA, Field DJ. 1997 Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* **37**, 3311–3325. (doi:10.1016/S0042-6989(97)00169-7)
 19. Laughlin SB, de Ruyter van Steveninck RR, Anderson JC. 1998 The metabolic cost of neural information. *Nat. Neurosci.* **1**, 36–41. (doi:10.1038/236)
 20. Sarpeshkar R. 1998 Analog versus digital: extrapolating from electronics to neurobiology. *Neural Comput.* **10**, 1601–1638. (doi:10.1162/089976698300017052)
 21. Balasubramanian V, Kimber D, Berry II MJ. 2001 Metabolically efficient information processing. *Neural Comput.* **13**, 799–815. (doi:10.1162/089976601300014358)
 22. Laughlin SB. 2001 Energy as a constraint on the coding and processing of sensory information. *Curr. Opin. Neurobiol.* **11**, 475–480. (doi:10.1016/S0959-4388(00)00237-3)
 23. Schreiber S, Machens CK, Herz AV, Laughlin SB. 2002 Energy-efficient coding with discrete stochastic events. *Neural Comput.* **14**, 1323–1346. (doi:10.1162/089976602753712963)
 24. Varshney LR, Sjöström PJ, Chklovskii DB. 2006 Optimal information storage in noisy synapses under resource constraints. *Neuron* **52**, 409–423. (doi:10.1016/j.neuron.2006.10.017)
 25. Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M. 1997 Adaptation of retinal processing to image contrast and spatial scale. *Nature* **386**, 69. (doi:10.1038/386069a0)
 26. Brenner N, Bialek W, de Ruyter Van Steveninck RR. 2000 Adaptive rescaling maximizes information transmission. *Neuron* **26**, 695–702. (doi:10.1016/S0896-6273(00)81205-2)
 27. Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. 2001 Efficiency and ambiguity in an adaptive neural code. *Nature* **412**, 787–792. (doi:10.1038/35090500)
 28. Hermundstad AM, Briguglio JJ, Conte MM, Victor JD, Balasubramanian V, Tkačik G. 2014 Variance predicts salience in central sensory processing. *eLife* **3**, e03722. (doi:10.7554/eLife.03722)
 29. Cover TM, Thomas JA. 1991 *Elements of information theory*. New York, NY: Wiley-Interscience.
 30. Sims CR. 2016 Rate–distortion theory and human perception. *Cognition* **152**, 181–198. (doi:10.1016/j.cognition.2016.03.020)
 31. Newsome WT, Britten KH, Anthony Movshon J. 1989 Neuronal correlates of a perceptual decision. *Nature* **341**, 52–54. (doi:10.1038/341052a0)
 32. Ernst MO, Bühlhoff HH. 2004 Merging the senses into a robust percept. *Trends Cogn. Sci.* **8**, 162–169. (doi:10.1016/j.tics.2004.02.002)
 33. Silvanto J, Soto D. 2012 Causal evidence for subliminal percept-to-memory interference in early visual cortex. *NeuroImage* **59**, 840–845. (doi:10.1016/j.neuroimage.2011.07.062)
 34. Shadlen MN, Newsome WT. 1994 Noise, neural codes and cortical organization. *Curr. Opin Neurobiol.* **4**, 569–579. (doi:10.1016/0959-4388(94)90059-0)
 35. Atwell D, Laughlin SB. 2001 An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21**, 1133–1145. (doi:10.1097/00004647-200110000-00001)
 36. Raichle ME, Gusnard DA. 2002 Appraising the brain's energy budget. *Proc. Natl Acad. Sci. USA* **99**, 10 237–10 239. (doi:10.1073/pnas.172399499)
 37. Hasenstaub A, Otte S, Callaway E, Sejnowski TJ. 2010 Metabolic cost as a unifying principle governing neuronal biophysics. *Proc. Natl Acad. Sci. USA* **107**, 12 329–12 334. (doi:10.1073/pnas.0914886107)
 38. Stevens CF. 2015 What the fly's nose tells the fly's brain. *Proc. Natl Acad. Sci. USA* **112**, 9460–9465. (doi:10.1073/pnas.1510103112)
 39. Marr D, Thomas Thach W. 1991 A theory of cerebellar cortex. In *From the retina to the neocortex*, pp. 11–50. Berlin, Germany: Springer.
 40. Burns JG, Foucaud J, Mery F. 2011 Costs of memory: lessons from 'mini' brains. *Proc. R. Soc. B* **278**, 923–929. (doi:10.1098/rspb.2010.2488)
 41. Marzen SE, DeDeo S. 2016 Weak universality in sensory tradeoffs. *Phys. Rev. E* **94**, 060101. (doi:10.1103/PhysRevE.94.060101)
 42. Tlusty T. 2008 Rate–distortion scenario for the emergence and evolution of noisy molecular codes. *Phys. Rev. Lett.* **100**, 048101. (doi:10.1103/PhysRevLett.100.048101)
 43. Herculano-Houzel S. 2011 Scaling of brain metabolism with a fixed energy budget per neuron: implications for neuronal activity, plasticity and evolution. *PLoS ONE* **6**, e17514. (doi:10.1371/journal.pone.0017514)
 44. Williams RW, Herrup K. 1988 The control of neuron number. *Annu. Rev. Neurosci.* **11**, 423–453. (doi:10.1146/annurev.ne.11.030188.002231)
 45. Berger T. 1971 *Rate–distortion theory*. Princeton, NJ: Prentice-Hall, Inc.
 46. Parrondo JMR, Horowitz JM, Sagawa T. 2015 Thermodynamics of information. *Nat. Phys.* **11**, 131–139. (doi:10.1038/nphys3230)
 47. Wolpert DH. 2016 The free energy requirements of biological organisms; implications for evolution. *Entropy* **18**, 138. (doi:10.3390/e18040138)
 48. Andrews BA, Iglesias PA. 2007 An information-theoretic characterization of the optimal gradient sensing response of cells. *Public Library Sci. Comput. Biol.* **3**, 1489–1497. (doi:10.1371/journal.pcbi.0030153)
 49. Palmer SE, Marre O, Berry MJ, Bialek W. 2015 Predictive information in a sensory population. *Proc. Natl Acad. Sci. USA* **112**, 6908–6913. (doi:10.1073/pnas.1506855112)
 50. Sims CR. 2015 The cost of misremembering: Inferring the loss function of visual working memory. *J. Vis.* **15**, 1–27. (doi:10.1167/15.3.2)
 51. Still S. 2009 Information-theoretic approach to interactive learning. *EuroPhys. Lett.* **85**, 28005. (doi:10.1209/0295-5075/85/28005)
 52. Tishby N, Polani D. 2011 Information theory of decisions and actions. In *Perception–action cycle*, pp. 601–636. Berlin, Germany: Springer.
 53. Tishby N, Pereira FC, Bialek W. 1999 The information bottleneck method. In *The 37th annual Allerton Conf. on Communication, Control, and Computing, 22–24 September, University of Illinois at Urbana-Champaign, Urbana, IL*. Urbana, IL: Coordinated Science Laboratory.
 54. Hinton GE. 1989 Connectionist learning procedures. *Artificial Intelligence* **40**, 185–234. (doi:10.1016/0004-3702(89)90049-0)
 55. Bäck T, Schwefel H-P. 1993 An overview of evolutionary algorithms for parameter optimization. *Evol. Comput.* **1**, 1–23. (doi:10.1162/evco.1993.1.1.1)
 56. Amari S-I. 1998 Natural gradient works efficiently in learning. *Neural Comput.* **10**, 251–276. (doi:10.1162/089976698300017746)
 57. Friston K. 2009 The free-energy principle: a rough guide to the brain? *Trends Cogn. Sci.* **13**, 293–301. (doi:10.1016/j.tics.2009.04.005)
 58. Frank SA. 2012 Wright's adaptive landscape versus Fisher's fundamental theorem. In *The adaptive landscape in evolutionary biology* (eds E Svensson, R Calsbeek). Oxford, UK: Oxford University Press.
 59. Marzen S, DeDeo S. 2017 PyRated: a Python package for rate distortion theory. Available at <https://bit.ly/pyrated> (accessed 5 March 2017).
 60. Shannon CE. 1959 Coding theorems for a discrete source with a fidelity criterion. *IRE Nat. Conv. Rec.* **4**, 142–163.
 61. Cooper VS, Lenski RE. 2000 The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature* **407**, 736–739. (doi:10.1038/35037572)
 62. Leiby N, Marx CJ. 2014 Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS Biol.* **12**, e1001789. (doi:10.1371/journal.pbio.1001789)
 63. West GB, Brown JH, Enquist BJ. 1997 A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126. (doi:10.1126/science.276.5309.122)

64. Castel AD, Pratt J, Drummond E. 2005 The effects of action video game experience on the time course of inhibition of return and the efficiency of visual search. *Acta Psychol.* **119**, 217–230. (doi:10.1016/j.actpsy.2005.02.004)
65. Feng J, Spence I, Pratt J. 2007 Playing an action video game reduces gender differences in spatial cognition. *Psychol. Sci.* **18**, 850–855. (doi:10.1111/j.1467-9280.2007.01990.x)
66. Morin-Moncet O, Therrien-Blanchet J-M, Ferland MC, Théoret H, West GL. 2016 Action video game playing is reflected in enhanced visuomotor performance and increased corticospinal excitability. *PLoS ONE* **11**, e0169013. (doi:10.1371/journal.pone.0169013)
67. Howard CJ, Wilding R, Guest D. 2016 Light video game play is associated with enhanced visual processing of rapid serial visual presentation targets. *Perception* **46**, 161–177. (doi:10.1177/0301006616672579)
68. Bongard JC. 2013 Evolutionary robotics. *Commun. ACM* **56**, 74–83. (doi:10.1145/2492007.2493883)
69. Lévi-Strauss C. 1969 *The elementary structures of kinship*. Boston, MA: Beacon Press.
70. Kikuchi DW, Pfennig DW. 2013 Imperfect mimicry and the limits of natural selection. *Q. Rev. Biol.* **88**, 297–315. (doi:10.1086/673758)
71. Edmunds M. 2000 Why are there good and poor mimics? *Biol. J. Linn. Soc.* **70**, 459–466. (doi:10.1111/j.1095-8312.2000.tb01234.x)
72. Strouse DJ, Schwab DJ. 2016 The deterministic information bottleneck. (<https://arxiv.org/abs/1604.00268>)