Physics of Living Systems

Information gain at the onset of habituation to repeated stimuli

Reviewed Preprint
v1 • October 25, 2024
Not revised

Giorgio Nicoletti, Matteo Bruzzone, Samir Suweis, Marco Dal Maschio, Daniel Maria Busiello

ECHO Laboratory, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland • Laboratory of Interdisciplinary Physics, Department of Physics and Astronomy "Galileo Galilei", University of Padova, Padova, Italy • Department of Mathematics "Tullio Levi-Civita", University of Padova, Padova, Italy • Department of Biomedical Science, University of Padova, Padova, Italy • Padova Neuroscience Center, University of Padova, Padova, Italy • Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

https://en.wikipedia.org/wiki/Open_access
Copyright information

eLife Assessment

This manuscript presents a **valuable** minimal model of habituation which is quantified by information theoretic measures. The results here could be of use in interpreting habituation behavior in a range of biological systems. However, the evidence presented is **incomplete** and would benefit from more rigorous approaches and a fuller accounting of the hallmarks of habituation.

https://doi.org/10.7554/eLife.99767.1.sa3

Abstract

Biological and living systems process information across spatiotemporal scales, exhibiting the hallmark ability to constantly modulate their behavior to ever-changing and complex environments. In the presence of repeated stimuli, a distinctive response is the progressive reduction of the activity at both sensory and molecular levels, known as habituation. Here, we solve a minimal microscopic model devoid of biological details to show that habituation is driven by negative feedback provided by a slow storage mechanism. Crucially, an intermediate level of habituation is associated with a steep increase in the information that the system collects on the external input over time. We find that the region characterized both by maximal information gain and by the onset of habituation can be retrieved if the system tunes its parameters to minimize dissipation and maximize information at the same time. We test our dynamical predictions against experimentally recorded neural responses in a zebrafish larva subjected to repeated looming stimulation. Our work makes a fundamental step towards uncovering the core mechanisms that shape habituation in biological systems, elucidating its information-theoretic and functional role.

Sensing, adaptation, and habituation mechanisms in biological systems span a wide range of temporal and spatial scales, from cellular to multi-cellular level, forming the basis for decision-making and the optimization of limited resources [1–8]. Prominent examples include the modulation of flagellar motion operated by bacteria according to changes in the local nutrient concentration [9–11], the regulation of immune responses through feedback mechanisms

[12, 13], the progressive reduction of neural activity in response to repeated looming stimulation [14, 15], and the maintenance of high sensitivity in varying environments for olfactory or visual sensing in mammalian neurons [16–20].

In the last decade, advances in experimental techniques fostered the quest for the core biochemical mechanisms governing information processing. Simultaneous recordings of hundreds of biological signals made it possible to infer distinctive features directly from data [21–24]. However, many of these approaches fall short of describing the connection between the underlying chemical processes and the observed behaviors [25–28]. To fill this gap, several works focused on the architecture of specific signaling networks, from tumor necrosis factor [12, 13] to chemotaxis [9, 29], highlighting the essential structural ingredients for their efficient functioning. An observation shared by most of these studies is the key role of a negative feedback mechanism to induce emergent adaptive responses [30–33]. Moreover, any information-processing system, biological or not, must obey information-thermodynamic laws that prescribe the necessity of a storage mechanism [34]. This is an unavoidable feature of numerous chemical signaling networks [9, 30] and biochemical realizations of Maxwell Demons [35, 36]. The storage of information consumes energy during processing [37, 38], and thus general sensing mechanisms have to take place out-ofequilibrium [3, 39–41]. Recently, the discovery of memory molecules [42–44] hinted at the implementation of storing mechanisms directly at the molecular scale. Overall, negative feedback, storage, and out-of-equilibrium conditions seem to be necessary requirements for a system to process environmental information and act accordingly. To quantify the performance of a biological informationprocessing system, theoretical developments made substantial progress in highlighting thermodynamics limitations and advantages [16, 45, 46], making a step towards linking information and dissipation from a molecular perspective [35, 47, 48].

Here, we consider an archetypal model for sensing that encapsulates all these key ingredients, i.e., negative feedback, storage, and energy dissipation, and study its response to repeated stimuli. Indeed, in the presence of dynamic environments, it is common for a biological system to keep encountering the same stimulus. Under these conditions, a progressive decay in the amplitude of the response is often observed, both at sensory and molecular levels. In general terms, such adaptive behavior is usually named *habituation* and it is a common phenomenon, from biochemical concentrations [49–51] to populations of neurons [14, 15, 52, 53]. In particular, habituation characterizes many neuronal circuits along the sensory-motor processing pathways in most living organisms, either invertebrates or vertebrates [52, 53]. While it has been proposed that inhibitory feedback mechanisms modulate the stimulus weight [15, 54], there are different hypotheses about the actual functional role of habituation in regulating the information flow, optimal processing, and sensitivity calibration [55], and controlling behavior and prediction during complex tasks [56–58]. Despite its ubiquity, the onset of habituation from general microscopic models remains unexplored, along with its functional advantages in terms of information gain and energy dissipation.

In this work, we tackle these questions. Our architecture is inspired by those found in real biological systems operating at different scales [12, 16] and resembles the topologies of minimal networks exhibiting adaptive features in different contexts [49, 50, 59]. By deriving the exact solution of this prototypical model, we identify that the key mechanism driving habituation is the negative feedback provided by slow information storage. We find that the information gain over time peaks at intermediate levels of habituation, uncovering that optimal processing performances are not necessarily tangled with maximal activity reduction. This optimal region can be retrieved by simultaneously minimizing dissipation and maximizing information, hinting at an a priori optimal region of operation for biological systems. Our results open the avenue to understanding the emergence of habituation, along with its information-theoretic advantage.

Results

Archetypal model for sensing in biological systems

We describe a model with three fundamental units: a receptor (R), a readout population (U), and a storage population (S) (**Figure 1a**). The presence of these three distinct components is a feature shared by several topologies exhibiting adaptive responses of various kinds [29, 49, 50, 59]. The role of the receptor is to sense external inputs, which we represent as a time-varying environment (H) described by the probability distribution $p_H(h, t)$. The receptor can be either active ($r = 1$) or passive ($r = 0$), with the two states separated by an energetic barrier, ΔE. A strong external signal favors activation of the receptor, while inhibition takes place through a negative feedback process mediated by the concentration of the storage, $[S]$. The negative feedback acts to reduce the level of activity of the system, and its effect on the receptor resembles known motifs found in biochemical systems (see **Figure 1b-e**) [12, 16]. We model the activation of the receptor by the environmental signal through a "sensing pathway" (superscript H). Instead, the inhibition mechanism affects an "internal pathway" of reactions (superscript I). By assuming that the rates follow an effective Arrhenius' law, we end up with:

$$\begin{aligned} \Gamma^{(H)}_{P\to A} &= e^{\beta(h-\Delta E)}\Gamma^0_R \qquad \Gamma^{(H)}_{A\to P} = \Gamma^0_R \\ \Gamma^{(I)}_{P\to A} &= e^{-\beta\Delta E}\Gamma^0_R \qquad \Gamma^{(I)}_{A\to P} = \Gamma^0_R e^{\beta\kappa[S]} \end{aligned} \tag{1}$$

where $\Gamma^0_R = \tau_R^{-1}$ sets the timescale of the receptor. For simplicity, the driving induced by inhibition appearing in $\Gamma^{(I)}_{A\to P}$ depends linearly on the concentration of S at a given time through a proportionality constant κ. Here, the inverse temperature β encodes the thermal noise, as lower values of β are associated with faster reactions (see Methods for a detailed discussion on model parameters). Crucially, the presence of two different transition pathways, motivated by molecular considerations and pivotal in many energy-consuming biochemical systems [35, 60, 61], creates an internal non-equilibrium cycle in receptor dynamics.

Whenever active, the receptor drives the production of the readout population U, which represents the direct response of the system to environmental signals. As such, we consider it to be the observable that characterizes habituation. It may describe, for example, photo-receptors or calcium concentration for olfactory or visual sensing mechanisms [14, 15, 17–20, 55]. We model its dynamics with a stochastic birth-and-death process:

$$\begin{aligned} \varnothing_U \xrightarrow{\Gamma_{u\to u+1}} U \qquad U \xrightarrow{\Gamma_{u+1\to u}} \varnothing_U \\ \Gamma_{u\to u+1} = e^{-\beta(V-cr)}\Gamma^0_U \qquad \Gamma_{u+1\to u} = (u+1)\Gamma^0_U \end{aligned} \tag{2}$$

where u denotes the number of molecules, $\Gamma^0_U = \tau_U^{-1}$ sets the timescale of readout production, and V is the energy needed to produce a readout unit. When the receptor is active, $r = 1$, this effective energetic cost is reduced by c by an effective additional driving. Thus, active receptors transduce the environmental energy into an active pumping on the readout node, allowing readout units to encode information on the external signal.

Finally, readout units stimulate the production of the storage population *S*. Its number of molecules s follows a controlled birth-and-death process [62–64]:

$$\begin{aligned} \varnothing_S \xrightarrow{\Gamma_{s\to s+1}(u)} S \qquad & S \xrightarrow{\Gamma_{s+1\to s}(u)} \varnothing_S \\ \Gamma_{s\to s+1}(u) = ue^{-\beta\sigma}\Gamma_S^0 \qquad & \Gamma_{s+1\to s} = (s+1)\Gamma_S^0 \end{aligned} \tag{3}$$

where σ is the energetic cost of a storage unit and Γ_S^0 sets the timescale, i.e., $\Gamma_S^0 = \tau_S^{-1}$. For simplicity, we set a first-order catalytic form for $\Gamma_{s\to s+1}(u)$ and allow for a maximum number of storage units, N_S, so that $[S] = s/N_S$. The storage may represent different molecular mechanisms at a coarse-grained level as, for example, memory molecules sensitive to calcium activity [42], synaptic depotentiation, and neural populations that regulate neuronal response [14, 15]. Storage units, as we will see, are responsible for encoding the readout response and play the role of a finite-time memory.

Our model, being devoid of specific biological details, can be declined to describe systems at very different scales (**Figure 1b-d**). We do not expect any detailed biochemical implementation to qualitatively change our results. However, we expect from previous studies [64] that the presence of multiple timescales in the system will be fundamental in shaping information between the different components. Thus, we employ the biologically plausible assumption that *U* undergoes the fastest evolution, while *S* and *H* are the slowest degrees of freedom [29, 65]. We have that $\tau_u \ll \tau_R \ll \tau_S \approx \tau_H$, where *TH* is the timescale of the environment.

The onset of habituation and its functional role

Habituation occurs when the response of the system, represented by the number of active readout units, decreases upon repeated stimulation. In our architecture, we expect it to emerge due to the increase in the storage population, which in turn provides an increasing negative feedback to the receptor. To study the onset and the features of habituation, we consider a switching exponential signal, $pH(h,t) \sim exp[-h/\langle H\rangle(t)]$. The time-dependent average $\langle H\rangle$ periodically switches between two values, $\langle H\rangle_{min}$ and $\langle H\rangle_{max}$, corresponding to a vanishing signal and strong stimulation of the receptor, respectively. Overall, the system dynamics is governed by four different operators, $\hat{W}_X$, with $X = R, U, S, H$, one for each population and one for the environment. The resulting master equation is:

$$\partial_t P = \left[\frac{\hat{W}_R(s,h)}{\tau_R} + \frac{\hat{W}_U(r)}{\tau_U} + \frac{\hat{W}_S(u)}{\tau_S} + \frac{\hat{W}_H}{\tau_H}\right] P\,, \tag{4}$$

where *P* denotes, in general, the joint propagator $P(u, r, s, h, t \,|\, u_0, r_0, s_0, h_0, t_0)$, with u_0, r_0, s_0 and h_0 initial conditions at time t_0. By taking advantage of the timescale separation, we can write an exact self-consistent solution to Eq. (4) at all times *t* (see Methods and Supplementary Information).

We assume that $\langle H\rangle$ switches to $\langle H\rangle_{max}$ at equally spaced intervals $t_1, \ldots, t_N$, each with the same duration ΔT. After a large number of inputs, the system reaches a time-periodic steady-state (see **Fig. 2d-e**). Thus, habituation is quantified by the change in the average response of the system:

$$\Delta\langle U\rangle = \langle U\rangle(t_\infty) - \langle U\rangle(t_1) \tag{5}$$

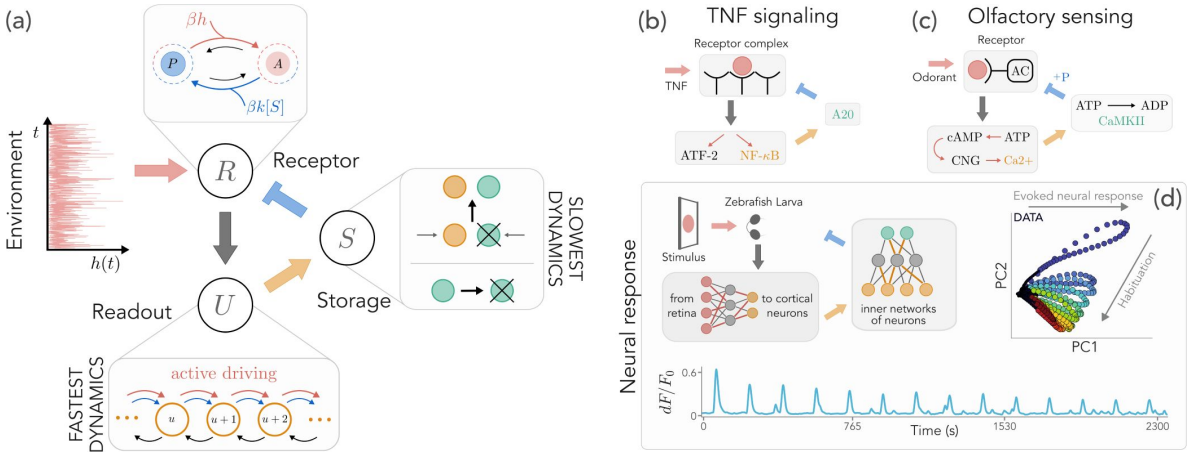


FIG. 1.

Sketch of the model architecture and biological examples at different scales. (a) A receptor *R* transitions between an active (*A*) and passive (*P*) state along two pathways, one used for sensing (red) and affected by the environment *h*, and the other (blue) modified by the storage concentration, [*S*]. An active receptor increases the response of a readout population *U* (orange), which in turn stimulates the production of storage units *S* (green) that provide negative feedback to the receptor. (b) In tumor necrosis factor (TNF) signaling, we can identify a similar architecture. The nuclear factor *NF* □ *κB* is produced after receptor binding to TNF. *NF* – *κB* modulates the encoding of the zinc-finger protein A20, which closes the feedback loop by inhibiting the receptor complex. (c) Similarly, in olfactory sensing, odorant binding induces the activation of adenylyl cyclase (AC). AC stimulates a calcium flux, eventually producing phosphorylase calmodulin kinase II (CAMKII) which phosphorylates and deactivates AC. (d) In neural response, multiple mechanisms take place at different scales. In zebrafish larvae, visual stimulation is pro jected along the visual stream from the retina to the cortex, a coarse-grained realization of the *R-U* dynamics. Neural habituation emerges upon repeated stimulation, as measured by calcium fluorescence signals (dF/F_0) and by the corresponding 2-dimensional PCA of the activity profiles.

where t_1 is the time of the first signal, and t_∞ is the time of a signal at the steady state. Whenever $\Delta \langle U \rangle < 0$, the system is habituating to the external inputs. In **Figure 2a**, we study habituation as a function of the inverse temperature δ and the energetic cost of storage, σ (see Methods). As expected, habituation is stronger at small σ, where a large storage production provides a strong negative feedback to the receptor, sharply decreasing $\langle U \rangle$.

During its dynamical evolution, the system encodes information on the environment H. We are particularly interested in how much information is captured by the readout population, which is measured by the mutual information between U and H at time t (see Methods):

$$I_{U,H}(t) = \mathcal{H}[p_U](t) - \int_0^\infty dh\, p_H(h,t)\mathcal{H}[p_{U|H}](t) \qquad (6)$$

where *H[p](t)* is the Shannon entropy of the probability distribution p, and $p_{U|H}$ denotes the conditional probability distribution of U given H. $I_{U,H}$ quantifies the system performance in terms of the information that the readout population captures on the external input at each time. Furthermore, it coincides with the entropy increase of the readout distribution:

$$k_B I_{U,H} = -k_B \left(\mathcal{H}[p_{U|H}] - \mathcal{H}[p_U]\right) = -\Delta \mathbb{S}_U \ . \qquad (7)$$

In **Figure 2b**, we show how the corresponding information gain $\Delta I_{U,H}$, defined in analogy to **Eq. (5)**, changes with β and σ. We find a region where the information gain is maximal. Surprisingly, this region corresponds to intermediate values of $\Delta \langle U \rangle$, suggesting that strong habituation driven by a low energetic cost of storage is ultimately detrimental to the system.

We can understand this feature by introducing the feedback information

$$\Delta I_f = I_{(U,S),H} - I_{U,H} > 0 \qquad (8)$$

which quantifies how much the simultaneous knowledge of U and S increases $I_{U,H}$ with respect to knowing solely U. We find that, during repeated external stimulation, the change in feedback information $\Delta\Delta I_f$, again in analogy to **Eq. (5)**, may be negative (**Figure 2c**). This indicates that the negative feedback on the receptor is impeding the information-theoretic performances of the system, independently of the habituation strength. Crucially, $\Delta\Delta I_f$ sharply increases in the region of maximal information gain, hinting that, at intermediate values of habituation, the information gain in the readout is driven by the storage mechanism. For the sake of simplicity, and to emphasize the information-theoretic advantage, we refer to this region of maximal information gain and intermediate habituation as the “onset” of habituation.

In **Figure 2(d-g)**, we show the evolution of the system for values of (β, σ) that lie in the region of maximal information gain. The readout activity decreases in time, modulating the response of the system to the repeated input (**Figure 2d**). This behavior resembles recent observations on habituation under analogous external conditions in various experimental systems [14, 15, 49–51]. We emphasize that the readout population is the fastest species at play, hence each point of the trajectory $\langle U \rangle$ (t) corresponds to a steady-state solution. As expected, the reduction of $\langle U \rangle$ is a direct consequence of the increase of the average storage population, $\langle S \rangle$ (**Figure 2e**). In this region, both the increase of $I_{U,H}$ and ΔI_f over time during habituation are optimal (**Figure 2f**). This behavior may seem surprising, since the increase in $I_{U,H}$ is concomitant to a reduction of the population that is encoding the signal. However, let us note that the mean of U is not directly related to the factorizability of the joint distribution $p_{U,H}$, i.e., to how much information on the signal is encoded in the readout. Furthermore, the inhibitory effect provided by S is enhanced by repeated stimuli, generating a stronger dependency between H and U over time.

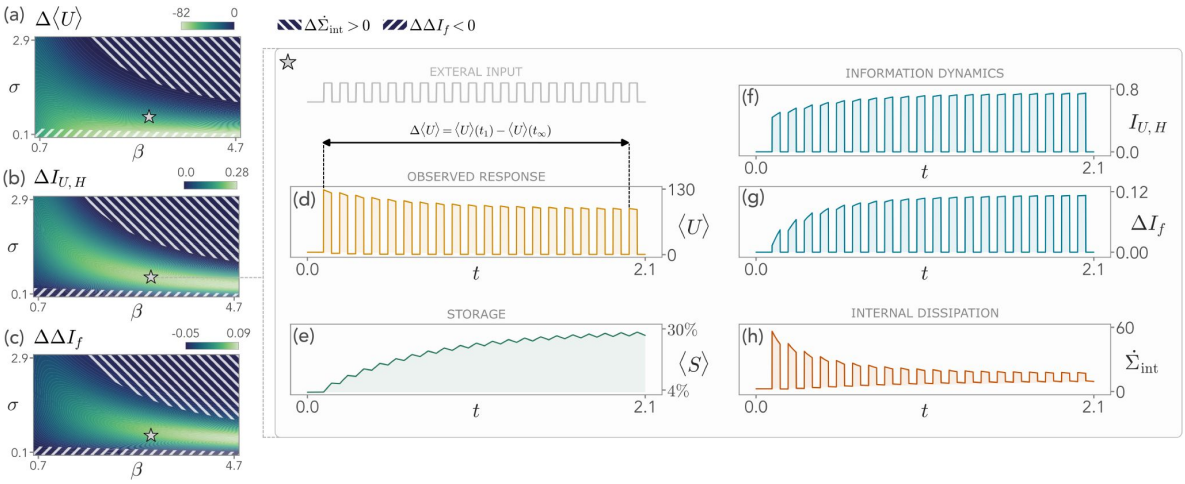


FIG. 2.

Evolution of the model under a switching external field *H(t)*. (a) The change in the average readout population Δ ⟨*U*⟩ between the first signal and after a large number of signals, as a function of the inverse temperature β and the energetic cost of storage σ. Δ ⟨*U*⟩ quantifies the habituation strength. (b) The change in the mutual information between the readout population and the external field, $\Delta I_{U,H}$. A region with maximal information gain corresponds to intermediate habituation strength. (c) The change in the feedback information ΔI_f indicates that, close to the region of maximal information gain, the storage favors information. (d-e) In the region of maximal information gain, the average number of readout units, ⟨*U*⟩, decreases with the number of repetitions, while the average storage concentration, ⟨*S*⟩, increases. At large times, the system reaches a periodic steady state. (f-g) In the same region, the information encoded on H through the readout, $I_{U,H}$, increases in time during habituation, boosting in turn the feedback information, ΔI_f. (h) The internal dissipation rate due to the production of *U* and *S*, $\dot{\Sigma}_{\rm int}$, decreases in time. Model parameters for panels (d-h) are $\beta = 3$, $\sigma = 0.6$ (in the unit measure of energy, for simplicity), and as specified in the Methods.

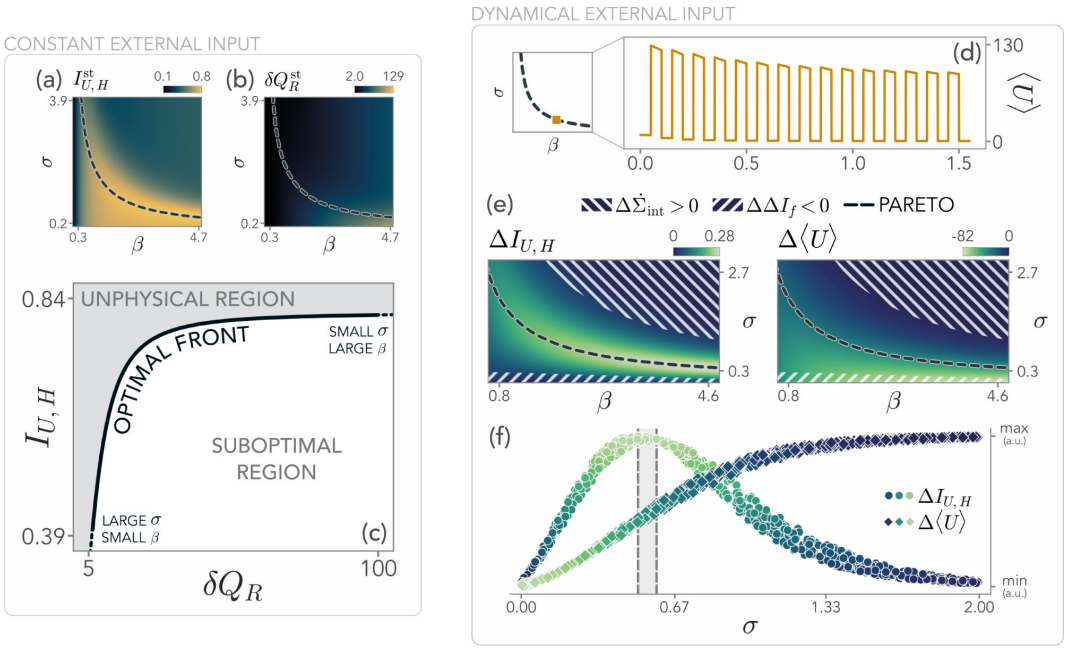


FIG. 3.

Optimality at the onset of habituation. (a-b) Contour plots in the (β, σ) plane of the stationary mutual information $I^{st}_{U,H}$ and the receptor dissipation per unit temperature, δQ^{st}_R, in the presence of a constant external input. (c) For a given value of β, the system can optimize σ to the Pareto front (black line) to simultaneously minimize δQ_R and maximize $I_{U,H}$. Each point in this space corresponds to a different strategy γ. If $\gamma = 0$, the system minimizes dissipation only, and if $\gamma = 1$ it only maximizes information. Below the front, the system exploits the available energy suboptimally, reaching lower values of information. In contrast, the region above the front is physically inaccessible. (d-e) In the presence of a dynamical input, the parameters defining the optimal front capture the region of maximal information gain and thus correspond to the onset of habituation, where $\langle U \rangle$ starts to be significantly smaller than zero. (f) Projection of $\Delta I_{U,H}$ and $\Delta \langle U \rangle$ along σ for a range of values of $\beta \in [3 - 3.5]$. The gray area enclosed by the dashed vertical lines indicates the location of the Pareto front for these values. $\Delta I_{U,H}$ clearly peaks at optimality, while $\Delta \langle U \rangle$ takes intermediate values.

The increase of $I_{U,H}$ comes along with another intriguing result. Since during habituation, the concentrations of the internal populations U and S change in time, we can quantify how much energy is required to support these processes. The rate of dissipation into the environment due to these internal mechanisms is (see Methods):

$$\dot{\Sigma}_{\rm int} = \sum_{u,s} \Big(\Gamma_{s\to s+1}\, p_{U,S}(u,s,t) + \\ -\Gamma_{s+1\to s}\, p_{U,S}(u,s+1,t) \Big) \log \frac{\Gamma_{s\to s+1}}{\Gamma_{s+1\to s}} \,. \tag{9}$$

We refer to $\dot{\Sigma}_{\rm int}$ as the internal dissipation of the system and, in **Figure 2h**, we show that it decreases over time, hinting at a synergistic thermodynamic advantage the onset of habituation.

Maximal information gain from an optimization principle

We now investigate whether the region of maximal information gain can be retrieved by means of an a priori optimization principle. To do so, we focus on the case of a constant environment. In this scenario, the system can tune its internal parameters to optimally respond to the statistics of an external input during a prolonged stimulation, i.e., the system "learns" the parameters while measuring an input with a large (infinite) observation time. Thus, the input statistics is given by $p^{\rm st}_H(h) \sim \exp[-h/\langle H\rangle_{\rm max}]$.

When the system reaches its steady state, the information that the readout has on the signal, $I^{\rm st}_{U,H}$, is estimated from the joint probability $p^{\rm st}_{U,S,R,H}$ (**Figure 3a**). At the same time, however, the system is consuming energy to maintain the receptor active. The receptor dissipation per unit temperature is given by

$$\delta Q_R = \left\langle \log\left(\frac{\Gamma^{(H)}_{P\to A}\Gamma^{(I)}_{A\to P}}{\Gamma^{(H)}_{A\to P}\Gamma^{(I)}_{P\to A}} \right) \right\rangle = \beta \left(\langle H\rangle + \kappa\sigma \frac{\langle S\rangle}{N_S} \right) \,,$$

as we show in **Figure 3b**. Large values of the mutual information compatible with minimal dissipation in the receptor can be obtained by maximizing the Pareto functional [66]:

$$\mathcal{L}(\beta,\sigma) = \gamma \frac{I_{U,H}(\beta,\sigma)}{\max(I_{U,H})} - (1-\gamma)\frac{\delta Q_R(\beta,\sigma)}{\max(\delta Q_R)} \tag{10}$$

where $\gamma \in [0,1]$ sets the strategy implemented by the system. If $\gamma \ll 1$, the system prioritizes minimizing dissipation, whereas if $\gamma \approx 1$ it acts to preferentially maximize information. The set of (β, σ) that maximize Eq. (10) defines a Pareto optimal front in the $(\delta Q_R, I_{U,H})$ space (**Figure 3c**). At fixed receptor dissipation, this front represents the maximum information between the readout and the external input that can be achieved. The region below the front is therefore suboptimal. Instead, the points ab ove the front are inaccessible, as higher values of $I_{U,H}$ cannot be attained without increasing δQ_R. We note that, since β usually cannot be directly controlled by the system, the Pareto front indicates the optimal a to which the system tunes at fixed β (see Methods and Supplementary Information for details).

Along this optimal front, we find that the system displays habituation (see **Figure 3d**). Furthermore, when plotted in the (β, σ) plane in the presence of a switching dynamical input, the front qualitatively corresponds to the region of maximal information gain and the onset of habituation, as we see in **Figure 3e**. Remarkably, this implies that once the system tunes its internal parameters to respond to a constant signal to maximize information and minimize

dissipation, it also learns to respond optimally to the time-varying input in terms of information gain. In **Figure 3f**, we show that at fixed β, the Pareto front (gray area) represents the region around the peak of $\Delta I_{U,H}$, where $\Delta \langle U \rangle$ attains intermediate values. In other words, the onset of habituation emerges spontaneously when the system attempts to activate its receptor as little as possible, while retaining information about the external environment.

The role of information storage

The presence of a storage mechanism is fundamental in our model. Furthermore, its role in mediating the negative feedback is suggested by several experimental and theoretical observations [9, 29–33]. Crucially, whenever the storage is eliminated from our model, habituation cannot take place, highlighting its key role in driving the observed dynamics.

In **Figure 4a**, we show that habituation and the change in the storage, $\Delta \langle S \rangle$, are deeply related to one another. The more <S> relaxes between two consecutive signals, the less the readout population reduces its activity. This ascribes to the storage population the role of an effective memory and highlights its dynamical importance for habituation. Moreover, the dependence of the storage dynamics on the interval between consecutive signals, ΔT, influences information gain as well. Indeed, increasing ΔT, we observe a decrease of the mutual information (**Figure 4b**) on the next stimulus, and the system needs to produce a larger number of readout units upon the new input. In the Supplementary Information, we further analyze the impact of different signal and pause durations.

We remark here that the proposed model is fully Markovian in its microscopic components, and the memory that governs readout habituation spontaneously emerges from the interplay among the internal timescales. In particular, recent works have highlighted that the storage needs to evolve on a slower timescale, comparable to that of the external input, in order to generate information in the receptor and in the readout [64]. In our model, instantaneous negative feedback implemented directly by U (bypassing the storage mechanism) leads to no time-dependent modulations (see Supplementary Information). Conversely, a readout population evolving on a timescale comparable to that of the signal cannot effectively mediate the negative feedback on the receptor since its population increase (as for the storage in the complete model) would not lead to habituation (see Supplementary Information). Thus, negative feedback has to be implemented by a separate degree of freedom evolving on a slow timescale.

Minimal features of neural habituation

In neural systems, habituation is typically measured as a progressive reduction of the stimulus-driven neuronal firing rate [14, 15, 52, 53, 55]. To test whether our minimal model can be used to capture the typical neural habituation dynamics, we measured the response of zebrafish larvae to repeated looming stimulations via volumetric multiphoton imaging [67]. From a whole-brain recording of ≈ 55000 neurons, we extracted a subpopulation of ≈ 2400 neurons in the optic tectum with a temporal activity profile that is most correlated with the stimulation protocol (see Methods).

Our model can be extended to qualitatively reproduce some features of the progressive decrease in neuronal response amplitudes. We identify each single readout with a subpopulation of binary neurons. A fraction of neurons are randomly turned on each time a readout unit is activated (see Methods). We tune the model parameters to have a comparable number of total active neurons at the first stimulus with respect to the experimental setting. Moreover, we set the pause and stimulus durations in line with the typical timescales of the looming stimulation. We choose the model parameters β and σ in such a way that the system operates close to the peak of information gain and the activity decrease over time is comparable to the activity decrease in experimental data (see Supplementary Information). In this way, we can focus on the effects of storage and feedback mechanisms without modeling further biological details. The patterns of the model-

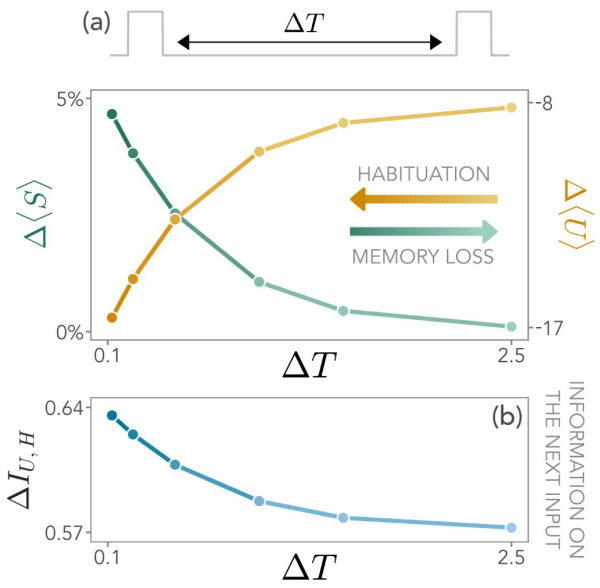


FIG. 4.

The role of memory in shaping habituation. (a) The system response depends on the waiting time ΔT_{pause} between two external signals. As ΔT_{pause} increases, the storage decays and thus memory is lost (green), and consequently the habituation of the readout population decreases (yellow). (b) As a consequence, the information $I_{U,H}$ that the system has on the field H when the new stimulus arrives decays as well. Model parameters for this figure are $\beta = 2.5$, $\sigma = 0.5$ in the unit measure of the energy, and as specified in the Methods.

generated activity are remarkably similar to the experimental ones (see **Figure 5a**). A 2dimensional embedding of the data via PCA (explained variance ≈ 70%) reveals that the evoked neural response is described by the first principal direction, while habituation is reflected in the second (**Figure 5b**). Remarkably, as we see in **Figure 5c**, this is the case in our minimal model as well, although the neural response is replaced by the switching on/off dynamics of the input.

Discussion

In this work, we considered a minimal architecture that serves as a microscopic and archetypal description of sensing processes across biological scales. Informed by theoretical and experimental observations, our model includes three fundamental mechanisms: a receptor, a readout population, and a storage mechanism that drives negative feedback. We have shown that our model ro-bustly displays habituation under repeated external inputs, a widespread phenomenon in both biochemical and neural systems. We find a regime of parameters of maximal information gain, where habituation drives an increase in the mutual information between external input and the system's response. Remarkably, the system can spontaneously tune to this region of parameters if it enforces an information-dissipation trade-off. In particular, optimal systems lie at the onset of habituation, characterized by intermediate levels of activity reduction, as both too-strong and too-weak negative feedback are detrimental to information gain. Our results suggest that the functional advantages of the onset of habituation are rooted in the interplay between energy dissipation and information gain.

Although minimal, our model can capture basic features of neural habituation, where it is generally accepted that inhibitory feedback mechanisms modulate the stimulus weight [54]. Remarkably, recent works reported the existence of a separate inhibitory neuronal population whose activity increases during habituation [15]. Our model suggests that this population might play the role of a storage mechanism, allowing the system to habituate to repeated signals. However, in neural systems, a prominent role in encoding both short- and long-term information is also played by synaptic plasticity [68, 69] as well as by memory molecules [42–44], at a biochemical level. A comprehensive analysis of how information is encoded and retrieved will most likely require all these mechanisms at once. Including an explicit connectivity structure with synaptic updates in our model may help in this direction, at the price of analytical tractability. Further, recent experiments also showed that by increasing the pause between two consecutive stimuli, the readout starts responding again, as theoretically predicted by our model [15]. Importantly, our framework allows us to formulate quantitative predictions of the system's response to subsequent stimulation. In particular, the increase in pause durations will decrease the habituation strength, until a typical time at which habituation should disappear. Comparison with experiments by modulating the frequency and intensity of stimulation will help identify the model parameters characterizing the system under investigation and, as such, its information-theoretic performance. Overall, these results hint at the fact that our minimal architecture may lay the foundation of habituation dynamics across vastly different biological scales.

Extensions of these ideas are manifold. Other *a priori* optimization principles for the system should be considered, focusing in particular on more detailed and realistic molecular schemes. Upon these premises, the possibility of inferring the underlying biochemical structure from observed behaviors is a fascinating direction [49]. Furthermore, since we focused on repetitions of statistically identical signals, it will be fundamental to characterize the system's response to diverse environments [70]. To this end, incorporating multiple receptors or storage populations may be needed to harvest information in complex conditions. In such scenarios, correlations between external signals may help reduce the encoding effort as, intuitively, S is acting as an information reservoir for the system. Moreover, such stored information might be used to make predictions on future stimuli and behavior, even if the detailed biological implementation of this complex task is still to be explored [56–58]. Indeed, living systems do not passively read

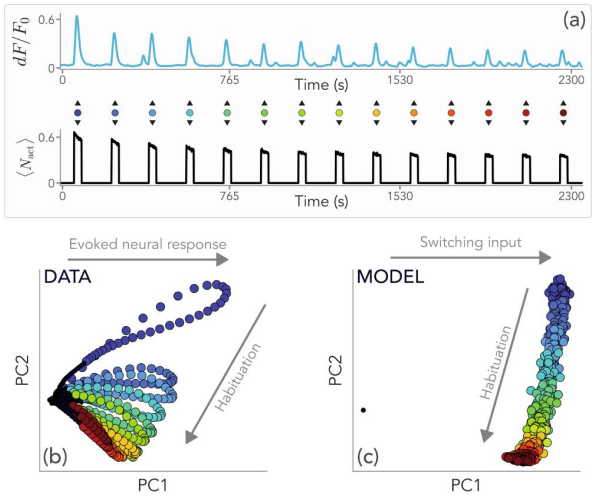


FIG. 5.

Habituation in zebrafish larvae. (a) Normalized neural activity profile in a zebrafish larva in response to the repeated presentation of visual stimulation, and comparison with the fraction of active neurons $\langle N_{act} \rangle = \langle N_{act} \rangle / N$ in our model with stochastic neural activation (see Methods). Stimuli are indicated with colored dots from blue to red as time increases. (b) PCA of experimental data reveals that habituation is captured mostly by the second principal direction, while features of the evoked neural response by the first one. Different colors indicate responses to different stimuli. (c) PCA of simulated neural activations. Although we cannot capture the dynamics of the evoked neural response with a switching input, the core features of habituation are correctly captured along the second principal direction. Model parameters are $\beta = 4.5$, $\sigma = 0.15$ in energy units, and as in the Methods, so that the system is tuned to the onset of habituation.

external signals but often act upon the environment. Both storage mechanisms and their associated negative feedback will remain core modeling ingredients, paving the way to understanding how this encoded information guides learning, predictions, and decision-making, a paramount question in many fields.

Our work serves as a fundamental framework for these ideas. On the one hand, it encapsulates key ingredients to support habituation while still being minimal enough to allow for analytical treatment. On the other hand, it may help the experimental quest for signatures of these physical ingredients in a variety of systems. Ultimately, our results show how habituation - a ubiquitous phenomenon taking place at strikingly different biological scales - may stem from an information-based advantage, shedding light on the optimization principle underlying its emergence and relevance for any biological system.

Acknowledgements

G.N., S.S., and D.M.B. acknowledge Amos Maritan for fruitful discussions. D.M.B. thanks Paolo De Los Rios for insightful comments. G.N. and D.M.B. acknowledge the Max Planck Institute for the Physics of Complex Systems for hosting G.N. during the initial stage of this work.

Methods

Model parameters.

The system is driven out of equilibrium by both the external field and the storage inhibition through the receptor dynamics, whose dissipation per unit temperature is δQ_R. The energetic barrier ($V - cr$) fixes the average values of the readout population both in the passive and active state, namely $\langle U \rangle_P = e^{-\beta V}$ and $\langle U \rangle_A = e^{-\beta(V-c)}$ (see Eq. (2)), and κ controls the effectiveness of the storage in inhibiting the receptor's activation. We assume that, on average, the activation rate due to the field is balanced by the feedback of a fraction $\alpha = \langle S \rangle / N_S$ of the storage population,

$$\left\langle \log \frac{\Gamma^{(H)}_{P\to A}}{\Gamma^{(I)}_{A\to P}} \right\rangle = \beta g \left(\langle H \rangle - \kappa \alpha \right) = 0 \quad \rightarrow \quad \kappa = \frac{\langle H \rangle}{\alpha} \, ,$$

so that we only need to fix α. Moreover, $\Delta E = 1$, $\langle U \rangle_A = 150$, $\langle U \rangle_P = 0.5$, $N_S = 25$, and $\alpha = 2/3$. We remark that the emerging features of the model are independent of the specific choice of these parameters. They affect the number of active units at each time step, but all the results presented here on the information gain during habituation remain valid. Furthermore, we typically consider the average of the exponentially distributed signal to be $\langle H \rangle_{\max} = 10$ and $\langle H \rangle_{\min} = 0.1$ (see Supplementary Information for details).

Overall, we are left with β and σ as free parameters. β quantifies the amount of thermal noise in the system, and at small β the thermal activation of the receptor hinders the effect of the field and makes the system almost unable to process information. Conversely, if β is high, the system must overcome large thermal inertia, increasing the dissipative cost. In this regime of weak thermal noise, we expect that, given a sufficient amount of energy, the system can effectively process information.

Timescale separation.

We solve our system in a timescale separation framework [64, 71, 72], where the storage evolves on a timescale that is much slower than all the other internal ones, i.e.,

$$\tau_U \ll \tau_R \ll \tau_S \approx \tau_H \ .$$

The fact that τ_S is the slowest timescale at play is crucial to making these components act as an information reservoir. This assumption is also compatible with biological examples. The main difficulty arises from the presence of the feedback, i.e. the field influences the receptor and thus the readout population, which in turn impacts the storage population and finally changes the deactivation rate of the receptor - schematically, $H \to R \to U \to S \to R$, but the causal order does not reflect the temporal one.

We start with the master equation for the propagator $P(u, r, s, h, t \,|\, u_0, r_0, s_0, h_0, t_0)$,

$$\partial_t P = \left[\frac{\hat{W}_U(r)}{\tau_U} + \frac{\hat{W}_R(s,h)}{\tau_R} + \frac{\hat{W}_S(u)}{\tau_S} + \frac{\hat{W}_H}{\tau_H} \right] P \ .$$

We rescale the time by τ_S and introduce two small parameters to control the timescale separation analysis, $\varepsilon = \tau_U/\tau_R$ and $\delta = \tau_R/\tau_H$. Since $\tau_S/\tau_H = O(1)$, we set it to 1 without loss of generality. We then write $P = P^{(0)} + \varepsilon P^{(1)}$ and expand the master equation to find $P^{(0)} = p^{\mathrm{st}}_{U|R}(u|r)\Pi$, with $\hat{W}_U \, p^{\mathrm{st}}_{U|R} = 0$. We obtain that Π obeys the following equation:

$$\partial_t \Pi = \left[\delta^{-1} \hat{W}_R(s,h) + \hat{W}_S(u) + \hat{W}_H \right] \Pi.$$

Yet again, $\Pi = \Pi^{(0)} + \delta\Pi^{(1)}$ allows us to write $\Pi^{(0)} = p^{\mathrm{st}}_{R|S,H}(r|s,h) F(s,h,t|s_0,h_0,t_0)$ at order $O(\delta^{-1})$, where $\hat{W}_R \, p^{\mathrm{st}}_{R|S,H} = 0$. Expanding first in ε and then in δ sets a hierarchy among timescales. Crucially, due to the feedback present in the system we cannot solve the next order explicitly to find F. Indeed, after a marginalization over r, we find $\partial_t F = \left[\hat{W}_H + \hat{W}_S(\bar{u}(s,h)) \right] F$, at order **O**(1), where $\bar{u}(s,h) = \sum_{u,r} u \, p^{\mathrm{st}}_{U|R}(u|r) p^{\mathrm{st}}_{R|S,H}(r|s,h)$. Hence, the evolution operator for F depends manifestly on s, and the equation cannot be self-consistently solved. To tackle the problem, we first discretize time, considering a small interval, i.e., $t = t_0 + \Delta t$ with $\Delta t \ll \tau_U$ and thus $\bar{u}(s,h) \approx u_0$. We thus find $F(s, h, t \,|\, s_0, h_0, t_0) = P(s,t \,|\, s_0, t_0) P_H(h, t \,|\, h_0, t_0)$ in the domain $t \in [t_0, t_0 + \Delta t]$, since H evolves independently from the system (see also Supplementary Information for analytical steps).

Iterating the procedure for multiple time steps, we end up with a recursive equation for the joint probability $pU, R, S, H\,(u, r, s, h, t_0 + \Delta t)$. We are interested in the following marginalization

$$p_{U,S}(u, t+\Delta t) = \sum_{r=0}^{1} \int_0^\infty dh \, p^{\mathrm{st}}_{U|R}(u|r) \, p^{\mathrm{st}}_{R|S,H}(r|h,s) \, p_H(h, t+\Delta t) \sum_{s'=0}^{N_S} \sum_{u'=0}^{\infty} P(s', t \to s, t+\Delta t)|u') p_{U,S}(u', s', t)$$

where $P(s',t \to s,t+\Delta t)$ is the propagator of the storage at fixed readout. This is the Chapman-Kolmogorov equation in the timescale separation approximation. Notice that this solution requires the knowledge of $p_{U,S}$ at the previous time-step and it has to be solved iteratively.

Explicit solution for the storage propagator.

To find a numerical solution to our system, we first need to compute the propagator $P(s_0, t_0 \to s,t)$. Formally, we have to solve the master equation

$$\begin{aligned}\partial_t P(s_0 \to s|u_0) = \Gamma_S^0 \Big[& e^{-\beta\sigma} u_0 P(s_0 \to s')\delta_{s',s-1} + \\ & + s' P(s_0 \to s')\delta_{s',s+1} + \\ & - P(s_0 \to s')\delta_{s',s} \left(s' + e^{-\beta\sigma} u_0\right) \Big]\end{aligned}$$

where we used the shorthand notation $P(s_0 \to s) = (s_0, t_0 \to s,t)$. Since our formula has to be iterated for small time-steps, i.e., $t - t_0 = \Delta t \ll 1$, we can write the propagator as follows

$$P(s_0, t_0 \to s, t_0 + \Delta t)|u_0) = p^{\mathrm{st}}_{S|U} + \sum_\nu w_\nu a^{(\nu)} e^{\lambda_\nu \Delta t}$$

where w_v and λ_v are respectively eigenvectors and eigenvalues of the transition matrix $\hat{W}_S(u_0)$,

$$\begin{aligned}\left(\hat{W}_S(u_0)\right)_{ij} &= e^{-\beta\sigma} u_0 && \text{if } i = j+1 \\ \left(\hat{W}_S(u_0)\right)_{ij} &= j && \text{if } i = j-1 \\ \left(\hat{W}_S(u_0)\right)_{ij} &= 0 && \text{otherwise}\end{aligned}$$

and the coeffiecients $a^{(v)}$ are such that

$$p_{S|U}(s_0, t_0 \to s, t_0 + \Delta t)|u_0) = p^{\mathrm{st}}_{S|U} + \sum_\nu w_\nu a^{(\nu)} = \delta_{s,s_0}.$$

Since eigenvalues and eigenvectors of $\hat{W}_S(u_0)$ might be computationally expensive to find, we employ another simplification. As $\Delta t \to 0$, we can restrict the matrix only to jumps to the n-th nearest neighbors of the initial state (s_0, t_0), assuming that all other states are left unchanged in small time intervals. We take $n = 2$ and check that the accuracy of this approximation against the full simulation for a limited number of time-steps.

Mean-field relationship.

We note that $\langle U \rangle$ and $\langle S \rangle$ satisfies the following mean-field relationship:

$$\frac{\langle U \rangle - \langle U \rangle_{r=1}}{\langle U \rangle_{r=1} - \langle U \rangle_{r=0}} = f_0\left(\frac{\langle S \rangle}{N_S}\right), \qquad \text{(S1)}$$

where $f_0(x)$ is an analytical function of its argument (see Supplementary Information). Eq. (S1) clearly states that only the fraction of active storage units is relevant to determining the habituation dynamics.

Mutual information.

Once we have $p_U(u, t)$ (obtained marginalizing $p_{U,S}$ over s) for a given $p_H(h, t)$, we can compute the mutual information

$$I_{U,H}(t) = \mathcal{H}[p_U](t) - \int_0^\infty dh\, p_H(h,t)\mathcal{H}[p_{U|H}](t)$$

where $\boldsymbol{H}$ is the Shannon entropy. For the sake of simplicity, we consider that the external field follows an exponential distribution $p_H(h, t) = \lambda(t)e^{-\lambda(t)h}$. Notice that, in order to determine such quantity, we need the conditional probability $p_{U|H}(u, t)$. In the Supplementary Information, we show how all the necessary joint and conditional probability distributions can be computed from the dynamical evolution derived above.

We also highlight here that the timescale separation implies $I_{S,H} = 0$, since

$$\begin{aligned} p_{S|H}(s,t|h) &= \sum_u p_{U,S|H}(u,s,t|h) \\ &= p_S(s,t)\sum_u\sum_r p^{\mathrm{st}}_{U|R}(u|r)p^{\mathrm{st}}_{R|S,H}(r|s,h) \\ &= p_S(s,t). \end{aligned}$$

Although it may seem surprising, this is a direct consequence of the fact that S is only influenced by H through the stationary state of U. Crucially, the presence of the feedback is still fundamental to promote habituation. Indeed, we can always write the mutual information between the field H and both the readout U and the storage S together as $I_{(U,S),H} = \Delta I_f + I_{U,H}$, where $\Delta I_f = I_{(U,S),H} - I_{U,H} = I_{(U,H),S} - I_{U,S}$. Since $\Delta I_f > 0$ (by standard information-theoretic inequalities), the storage is increasing the information of the two populations together on the external field. Overall, although S and H are independent in this limit, the feedback is paramount in shaping how the system responds to the external field and stores information about it.

Dissipation of internal processes

The production of readout, u, and storage, s, molecules requires energy. From the modeling of their dynamics as controlled stochastic birth-and-death processes, we quantify the dissipation into the environment using the environmental contribution of the Schnakenberg entropy production, which is also the only one that survives at stationarity [73]. We have:

$$\begin{aligned} \dot{\Sigma}_{\mathrm{int}} &= \sum_{u,s}\big(\Gamma_{u\to u+1}p_{U,S}(u,s,h,t) + \\ &\quad -\Gamma_{u+1\to u}p_{U,S}(u+1,s,h,t)\big)\log\frac{\Gamma_{u\to u+1}}{\Gamma_{u+1\to u}} + \\ &+ \sum_{u,s}\big(\Gamma_{s\to s+1}p_{U,S}(u,s,h,t) + \\ &\quad -\Gamma_{s+1\to s}p_{U,S}(u,s+1,h,t)\big)\log\frac{\Gamma_{s\to s+1}}{\Gamma_{s+1\to s}} \end{aligned}$$

where we indicated all possible dependencies in the joint probability distribution. By employing the timescale separation [71], and noting that $\Gamma_{u\to u\pm1}$ do not depend on s, we finally have:

$$\dot{\Sigma}_{\rm int} = \sum_{u,s} (\Gamma_{s\to s+1} p_{U,S}(u,s,h,t) + \\ -\Gamma_{s+1\to s} p_{U,S}(u,s+1,h,t)) \log \frac{\Gamma_{s\to s+1}}{\Gamma_{s+1\to s}}$$

As this quantity decreases during habituation, the system tends to dissipate less and less into the environment to produce the internal populations that are required to encode and store the external signal.

Pareto optimization.

We perform a Pareto optimization at stationarity in the presence of a prolonged stimulation. We seek the optimal values of (β, σ) by maximizing the functional

$$\mathcal{L}(\beta,\sigma) = \gamma \frac{I_{U,H}(\beta,\sigma)}{\max(I_{U,H})} - (1-\gamma)\frac{\delta Q_R(\beta,\sigma)}{\max(\delta Q_R)}$$

where $\gamma \in [0,1]$. Hence, we maximize the information between the readout and the field, simultaneously minimizing the dissipation of the receptor induced by both the signal and feedback process, as discussed in the main text. The values are normalized since, in principle, they can span different orders of magnitudes. In the Supplementary Information, we detailed the derivation of the Pareto front and show that qualitatively similar results can be obtained for a 3-d Pareto-like surface obtained by maximizing also the feedback information, ΔI_f.

Recording of whole brain neuronal activity in zebrafish larvae.

Acquisitions of the zebrafish brain activity were carried out in one Elavl3:H2BGCaMP6s larvae at 5 days post fertilization raised at 28°C on a 12 h light/12 h dark cycle according to the approval by the Ethical Committee of the University of Padua (61/2020 dal Maschio). The subject was embedded in 2 percent agarose gel and brain activity was recorded using a multiphoton system with a custom 3D volumetric acquisition module. Data were acquired at 30 frames per second covering an effective field of view of about 450 × 900um with a resolution of 512 × 1024 pixels. The volumetric module acquires a volume of about 180 – 200um in thickness encompassing 30 planes separated by about 7um, at a rate of 1 volume per second, sufficient to track the slow dynamics associated with the fluorescence-based activity reporter GCaMP6s. Visual stimulation was presented in the form of a looming stimulus with 150s intervals, centered with the fish eye (see Supplementary Information). Neurons identification and anatomical registrations were performed as described in [67].

Data analysis.

The acquired temporal series were first processed using an automatic pipeline, including motion artifact correction, temporal filtering with a 3s rectangular window, and automatic segmentation. The obtained dataset was manually curated to resolve segmentation errors or to integrate cells not detected automatically. We fit the activity profiles of about 55000 cells with a linear regression model using a set of base functions representing the expected responses to each stimulation event. These base functions have been obtained by convolving the exponentially decaying kernel of the GCaMP signal lifetime with square waveforms characterizing the presentation of the corresponding visual stimulus. The resulting score coefficients of the fit were used to extract the cells whose score fell within the top 5% of the distribution, resulting in a population of ≈ 2400

neurons whose temporal activity profile correlates most with the stimulation protocol. The resulting fluorescence signals $F^{(i)}$ were processed by removing a moving baseline to account for baseline drifting and fast oscillatory noise [74]. See Supplementary Information.

Model for neural activity.

Here, we describe how our framework is modified to mimic neural activity. Each readout unit, u, is interpreted as a population of N neurons, i.e., a region dedicated to the sensing of a specific input. Storage can be implemented, for example, as an inhibitory neural population, in line with recent observations in [15]. When a readout population is activated at time t, each of its N neurons fires with a probability p. We set $N = 20$ and $p = 0.5$. N has been set to have the same number of observed neurons in data and simulations, while p only controls the dispersal of the points in **Fig. 5c**, thus not altering the main message. The dynamics of each readout unit follows our dynamical model. Due to habituation, some of the readout units activated by the first stimulus will not be activated by subsequent stimuli. Although the evoked neural response cannot be captured by this extremely simple model, its archetypal ingredients (dissipation, storage, and feedback) are informative enough to reproduce the low-dimensional habituation dynamics found in experimental data.

S1. Detailed Solution of the Master Equation

Consider the transition rates introduced in the main text:

$$\begin{aligned} \Gamma^{(H)}_{P\to A} &= e^{\beta(h-\Delta E)}\Gamma^0_H & \Gamma^{(H)}_{A\to P} &= \Gamma^0_H \\ \Gamma^{(I)}_{P\to A} &= e^{-\beta\Delta E}\Gamma^0_I & \Gamma^{(I)}_{A\to P} &= \Gamma^0_I e^{\beta\kappa s/N_s} \\ \Gamma_{u\to u+1} &= e^{-\beta(V-cr)}\Gamma^0_U & \Gamma_{u+1\to u} &= (u+1)\Gamma^0_U \\ \Gamma_{s\to s+1} &= e^{-\beta\sigma}u\Gamma^0_S & \Gamma_{s+1\to s} &= (s+1)\Gamma^0_S. \end{aligned}$$

We set a reflective boundary for the storage at $s = N_S$, corresponding to the maximum amount of storage molecules in the system. Moreover, for the sake of simplicity, we take $\Gamma^0_I = \Gamma^0_H \equiv \Gamma^0_R$. Retracing the steps of the Methods, the Master Equation governing the evolution of the propagator of all variables, $P(u, r, s, h, t\,|\,u_0, r_0, s_0, h_0, t_0)$, is:

$$\partial_t P = \left[\frac{\hat{W}_U(r)}{\tau_U} + \frac{\hat{W}_R(s,h)}{\tau_R} + \frac{\hat{W}_S(u)}{\tau_S} + \frac{\hat{W}_H}{\tau_H}\right] P\,. \tag{S2}$$

We solve this equation employing a timescale separation, i.e., $\tau_U \ll \tau_R \ll \tau_S \sim \tau_H$, where $\tau_X = \Gamma^0_X$ for $X = U, R, S$ and τ_H is the typical timescale of the signal dynamics. Motivated by several biological examples, we assumed that the readout population undergoes the fastest dynamics, while storage and signal evolution are the slowest ones. Defining $\varepsilon = \tau_U/\tau_R$ and $\delta = \tau_R/\tau_H$, and setting $\tau_S/\tau_H = 1$ without loss of generality, we have:

$$\partial_t P = \left[\epsilon^{-1}\delta^{-1}\hat{W}_U(r) + \delta^{-1}\hat{W}_R(s,h) + \hat{W}_S(u) + \frac{\tau_S}{\tau_H}\hat{W}_H\right] P\,. \tag{S3}$$

We propose a solution in the following form, $P = P^{(0)} + \varepsilon P^{(1)}$. By inserting this expression in the equation above, and solving order by order in ε, at order ε^{-1}, we have that:

$$P^{(0)} = p^{\mathrm{st}}_{U|R}(u|r)\Pi(r,h,t|r_0,h_0,t_0) \tag{S4}$$

where p^{st} solves the master equation for the readout evolution at a fixed r:

$$0 = p^{\mathrm{st}}_{U|R}(u+1)[u+1] + p^{\mathrm{st}}_{U|R}\alpha(r) - p^{\mathrm{st}}_{U|R}(u)[u+\alpha(r)] \tag{S5}$$

with $\alpha(r) = e^{-\beta(V-cr)}$. Hence,

$$p^{\mathrm{st}}_{U|R}(u|r) = e^{-\alpha(r)}\frac{\alpha(r)^u}{u!}. \tag{S6}$$

At order ε^0, we find the equation for Π, also reported in the Methods:

$$\partial_t \Pi(r,h,t|r_0,h_0,t_0) = \left[\delta^{-1}\hat{W}_R(h) + \hat{W}_S(u) + \hat{W}_H\right]\Pi(r,h,t|r_0,h_0,t_0)\ . \tag{S7}$$

To solve this equation, we propose a solution of the form $\Pi = \Pi^{(0)} + \delta\Pi^{(1)}$. Hence, again, at order δ^{-1}, we have that $\Pi^{(0)} = p^{\mathrm{st}}_{R|S,H}(r|s,h)\hat{F}(s,h,t|s_0,h_0,t_0)$, where $p^{\mathrm{st}}_{R|S,H}$ satisfy the steady-state equation for the fastest degree of freedom, with all the others fixed. In the case, it is just the solution of the rate equation for the receptor:

$$p^{\mathrm{st}}_{R|H,S}(r=1) = \frac{\Gamma^{\mathrm{eff}}_{P\to A}}{\Gamma^{\mathrm{eff}}_{P\to A} + \Gamma^{\mathrm{eff}}_{A\to P}}, \quad p^{\mathrm{st}}_{R|H}(r=0) = 1 - p^{\mathrm{st}}_{R|H}(r=1,t) \tag{S8}$$

where $\Gamma^{\mathrm{eff}}_{P\to A} = \Gamma^{(I)}_{P\to A} + \Gamma^{(H)}_{P\to A}$, and the same for the reverse reaction. At order δ^{-1}, we have an equation for F:

$$\partial_t F(s,h,t|s_0,h_0,t_0) = \sum_{r,u} p^{\mathrm{st}}_{U|R}(u|r)\left[\hat{W}_S(u) + \hat{W}_H\right]\left[p^{\mathrm{st}}_{R|S,H}(r|s,h)F(s,h,t|s_0,h_0,t_0)\right] \tag{S9}$$

As already explained in the Methods, due to the feedback, this equation cannot be solved explicitly. Indeed, the operator governing the evolution of F is:

$$\hat{W}_{\mathrm{eff}} = \hat{W}_H + \sum_u p^{\mathrm{st}}_{U|S,H}(u|s,h)\hat{W}_S(u) = \hat{W}_H + \hat{W}_S\left(\sum_u u\, p^{\mathrm{st}}_{U|S,H}(u|s,h)\right) = \hat{W}_H + \hat{W}_S\left(\bar{u}(s,h)\right) \tag{S10}$$

with $p^{\mathrm{st}}_{U|S,H}(u|s,h) = \sum_r p^{\mathrm{st}}_{U|R}(u|r)p^{\mathrm{st}}_{R|S,H}(r|s,h)$ and using the linearity of $\hat{W}_S(u)$. In order to solve this equation, we shall assume that $\bar{u}(s,h) = u_0$, bearing in mind that this approximation holds if t is small enough, i.e., $t = t_0 + \Delta t$ with $\Delta t \ll \tau_u$. Therefore, for a small interval, we have:

$$\partial_t F(s,h,t_0+\Delta t|s_0,h_0,t_0) = \left[\hat{W}_S(u_0) + \hat{W}_H\right]F(s,h,t|s_0,h_0,t_0) \tag{S11}$$

Overall, we end up with the following joint probability of the model at time $t_0 + \Delta t$:

$$p_{U,R,S,H}(u,r,s,h,t_0+\Delta t) = \\ = \sum_{u_0,s_0} p^{\mathrm{st}}_{U|R}(u|r) p^{\mathrm{st}}_{R|S,H}(r|s,h) P(s,t_0+\Delta t|s_0,u_0,t_0) \int dh_0 P_H(h,t_0+\Delta t|h_0,t_0) p_{U,S,H}(u_0,s_0,h_0,t_0) \\ = \sum_{u_0,s_0} p^{\mathrm{st}}_{U|R}(u|r) p^{\mathrm{st}}_{R|S,H}(r|s,h) P(s,t_0+\Delta t|s_0,u_0,t_0) p_{U,S}(u_0,s_0,t_0) p_H(h,t_0+\Delta t) \tag{S12}$$

where $\int dh_0 P_H(h, t_0 + \Delta t \,|\, h_0, t_0) p_{U,S,H}(u_0, s_0, h_0, t_0) = p_{U,S}(u_0, s_0, t_0) pH(h, t_0 + \Delta t)$ since H at time $t_0 + \Delta t$ is independent of S and U. When propagating the evolution through intervals of duration Δt, we also assume that H evolves independently since it is an external variable, while affecting the evolution of the other degrees of freedom. This structure reflects into the equation above. For simplicity, we prescribe $p_H(h, t)$ to be an exponential distribution, $p_H(h, t) = \lambda(t) e^{-\lambda(t)h}$, and solve iteratively **Eq. (S12)** from t_0 to a given T in steps of duration Δt, as indicated above. This complex iterative solution arises from the timescale separation because of the cyclic feedback structure: $\{S, H\} \to R \to U \to S$. This solution corresponds explicitly to

$$p_{U,S}(u,t+\Delta t) = \sum_{r=0}^{1} \int_0^\infty dh\, p^{\mathrm{st}}_{U|R}(u|r)\, p^{\mathrm{st}}_{R|S,H}(r|h,s)\, p_H(h,t+\Delta t) \sum_{s'=0}^{N_S} \sum_{u'=0}^{\infty} P(s',t\to s,t+\Delta t|u') p_{U,S}(u',s',t) \tag{S13}$$

where $P(s', t \to s, t + \Delta t)$ is the propagator of the storage at fixed readout. This is the Chapman-Kolmogorov equation in the time-scale separation approximation. Notice that this solution requires the knowledge of $p_{U,S}$ at the previous time-step and it has to be solved iteratively. Both p_U and p_S can be obtained by an immediate marginalization.

As detailed in the Methods, the propagator $P(s_0, t_0 \to s, t)$, when restricted to small time intervals, can be obtained by solving the birth-and-death process for storage molecules at fixed readout, limiting the state space only to n nearest neighbors (we checked that our results are robust increasing n for the selected simulation time step).

S2. Information-Theoretic Quantities

By direct marginalization of **Eq. (S13)**, we obtain the evolution of $p_U(u, t)$ and $p_S(s, t)$ for a given $p_H(h, t)$. Hence, we can compute the mutual information as follows:

$$I_{U,H}(t) = \mathcal{H}[p_U](t) - \int_0^\infty dh\, p_H(h,t) \mathcal{H}[p_{U|H}](t) = -\frac{\Delta \mathbb{S}_U}{k_B} \tag{S14}$$

where $\boldsymbol{H}[p_X]$ is the Shannon entropy of X, and $\Delta \mathbb{S}_U$ is the reduction in the entropy of U due to repeated measurement (see main text). Notice that, in order to determine such quantity, we need the conditional probability $p_{U|H}(u, t)$. This distribution represent the probability that, at a given time, the system jumps at a value u in the presence of a given field h. In order to compute it, we can write

$$p_{U|H}(u,t+\Delta t) = \sum_{s=0}^{M_S} \sum_{r=0}^{1} p^{\mathrm{st}}_{U|R}(u|r)\, p^{\mathrm{st}}_{R|S,H}(r|h,s)\, p_S(s,t+\Delta t) \tag{S15}$$

by definition. The only dependence on h enters in $p^{\mathrm{st}}_{R|S,H}$ through the $e^{\beta h}$ dependence in the rates.

Analogously, all the other mutual information can be obtained. Notably, as we showed in the Methods, $I_{S,H} = 0$, as a consequence of the time-scale separation. Crucially, the presence of the feedback is still fundamental to effectively process information about the signal. This effect can be quantified through $\Delta I_f = I_{(U,S),H} - I_{U,H} > 0$, which we name feedback information, as it captures how much the knowledge of S and U together helps encoding information about the signal with respect to U alone. In terms of system entropy, we equivalently have:

$$k_B \Delta I_f = -\Delta \mathbb{S}_{U,S} + \Delta \mathbb{S}_U > 0 \tag{S16}$$

that highlights how much the effect of S (feedback) reduces the entropy of the system due to repeated measurements.

Practically speaking, in order to evaluate $I_{(U,S),H}$, we exploit the following equality:

$$I_{(U,S),H} = \mathcal{H}\left[p_{U,S}\right](t) - \int_0^\infty dh p_H(h,t) \mathcal{H}\left[p_{U,S|H}\right](t). \tag{S17}$$

for which we need $p_{U,S|H}$, that can be found noting that

$$p_{U,S}(u,s,t) = p_{U|S}(u,t|s)p_S(s,t) = \int dh \sum_r p^{\mathrm{st}}_{U|R}(u|r) p^{\mathrm{st}}_{R|S,H}(r|s,h) p_S(s,t) p_H(h,t) \tag{S18}$$

from which we immediately see that

$$p_{U,S|H}(u,s,t) = \sum_{r=0}^{1} p^{\mathrm{st}}_{U|R}(u|r)\, p^{\mathrm{st}}_{R|S,H}(r|h,s) p_S(s,t) \tag{S19}$$

which we can easily computed at any given time t.

S3. Mean-Field Relation Between Average Readout and Storage

Fixing all model parameters, the average value of storage, $\langle S \rangle$, and readout, $\langle U \rangle$, is numerically determined by solving iteratively the system, as shown above. However, an analytical relation between these two quantities can be found starting from the definition of $\langle U \rangle$:

$$\langle U \rangle = \sum_{u,s} u P^{\mathrm{st}}_{U,S}(u,s) = \sum_{u,s} u P^{\mathrm{st}}_{U|S}(u|s) P^{\mathrm{st}}_S(s) = \sum_{u,s,r} u P^{\mathrm{st}}_{U|R}(u|r) P^{\mathrm{st}}_{R|S}(r|s) P^{\mathrm{st}}_S(s) \tag{S20}$$

Then, inserting the expression for the stationary probability that we know analytically:

$$\langle U \rangle = \sum_s \left(\left\langle u P^{\mathrm{st}}_{U|R}(u|r=0) \right\rangle P^{\mathrm{st}}_{R|S}(r=0|s) + \left\langle u P^{\mathrm{st}}_{U|R}(u|r=1) \right\rangle P^{\mathrm{st}}_{R|S}(r=1|s) \right) P^{\mathrm{st}}_S(s) \tag{S21}$$

where $P^{\rm st}_{R|S} = \int dh P^{\rm st}_{R|H,S} P_H dh \equiv f_R(\rho_S)$ has a complicated expression involving the hypergeometric function ${}_2F_1$ in terms of model parameters and only the concentration of S, $\rho S = s/N_S$ (the explicit derivation of this formula is not shown here). Then, we have:

$$\langle U \rangle = \sum_s \left(e^{-\beta V} f_0(\rho_S) + e^{-\beta(V-c)}(1 - f_0(\rho_S)) \right) P^{\rm st}_S(s) \tag{S22}$$

Since we do not have an analytical expression for $P^{\rm st}_S(s)$, we employ the mean-field approximation, reducing all the correlation functions to products of averages:

$$\langle U \rangle = e^{-\beta(V-c)} + e^{-\beta V} f_0(\bar{\rho}_S) \left(1 - e^{\beta c}\right) \tag{S23}$$

where $\bar{\rho}_S = \langle S \rangle / N_S$. This clearly shows that, given a set of model parameters, $\langle U \rangle$ and the average concentration of storage, $\bar{\rho}_S$ are related. In particular, introducing the change of parameters presented in the Methods, we have the following collapse:

$$\frac{\langle U \rangle - \langle U \rangle_A}{\langle U \rangle_A - \langle U \rangle_P} = f_0(\bar{\rho}_S) \tag{S24}$$

where $\langle U \rangle_A$ and $\langle U \rangle_P$ are respectively the average of U fixing $r = 1$ (active receptor) and $r = 0$ (passive receptor). It is also possible to perform an expansion of f_0 which numerically results to be very precise:

$$\frac{\langle U \rangle - \langle U \rangle_A}{\langle U \rangle_A - \langle U \rangle_P} = \frac{a_{-1}(\lambda_H, \beta, g)}{z} + a_0(\lambda_H, \beta) + a_1(\lambda_H, \beta, g) z^{-1-\lambda_H/\beta} + a_2(\lambda_H, \beta) z^{-\lambda_H/\beta} \tag{S25}$$

where $z = e^{\beta \Delta E}\left(1 + g\, e^{\beta \bar{\rho}_S / \alpha \lambda_H}\right)$. Since all these relations just depend on the average concentration of the storage, it is natural to ask what happens when $N_S \to N'_S = n N_S$. Fixing all the remaining parameters, both $\langle U \rangle$ and $\bar{\rho}_S$ will change, still satisfying the mutual relation presented above. Let us consider, for N'_S, the stationary solution that has the same concentration of S, i.e., $(\bar{\rho}_S)_{N'_S} = (\bar{\rho}_S)_{N_S}$. As a consequence of the scaling relation, $\langle U \rangle_{N'_S} \neq \langle U \rangle_{N_S}$. Considering $\langle U \rangle_P \approx 0$ in both settings, we can ask ourselves what is the factor γ such that $\gamma(\langle U \rangle_A)_{N_S} = (\langle U \rangle_A)_{N'_S}$. Since u only enters linearly in the dynamics of the storage, and the mutual relation only depends on the concentration of S, we guess that $\gamma = 1/n$, as numerically observed. As stated in the main text, we can finally conclude that the storage concentration is the most relevant quantity in our model to determine the effect of the feedback and characterize the dynamical evolution. This observation makes our conclusions more robust, as they do not depend on the specific choice for the storage reservoir since there always exists a scaling relation connecting $\langle U \rangle$ and $\bar{\rho}_S$. As such, changing the value of the model parameters we fixed, will only affect the number of active molecules without modifying the main results presented in this work.

S4. The Necessity of Storage

Here, we discuss in detail the necessity of slow storage implementing the negative feedback to have habituation. We will first investigate the possibility that negative feedback, necessary for any kind of habituative behaviors, is implemented directly through the readout population that

undergoes a fast dynamics. We will analytically show that this limit leads to the absence of habituation, hinting at the necessity of having a slow dynamical feedback in the system (Sec. S4 1). Then, we will study the system in the scenario in which U applies the feedback, bypassing the storage S, but it acts as a slow variable. Solving the Master Equation through our iterative numerical method, we show that, also in this case, habituation disappears (Sec. S4 2). These results suggest that not only the feedback must be applied by a slow variable, but that such a slow variable must have a role different from the readout population, in line with recent observations in neural systems [15]. The model proposed in the main text is indeed minimal in this respect, other than compatible with biological examples.

1. Dynamical feedback cannot be implemented by a fast readout

If the storage is directly implemented by the readout population, the transition rates get modified as follows:

$$\begin{aligned} \Gamma^{(H)}_{P\to A} &= e^{\beta(h-\Delta E)}\Gamma^0_R \qquad & \Gamma^{(H)}_{A\to P} &= \Gamma^0_R \\ \Gamma^{(C)}_{P\to A} &= e^{-\beta\Delta E}\Gamma^0_R \qquad & \Gamma^{(C)}_{A\to P} &= e^{\beta\theta u}\Gamma^0_R \\ \Gamma_{u\to u+1} &= e^{-\beta(V-cr)}\Gamma^0_U \qquad & \Gamma_{u+1\to u} &= \Gamma^0_U(u+1) \end{aligned} \tag{S26}$$

At this level, θ is a free parameter playing the same role as κ/N_s in the complete model with the storage. We start again from the master equation for the propagator $P(u, r, h, t\,|\,u_0, r_0, h_0, t_0)$:

$$\partial_t P = \left[\frac{\hat{W}_U(r)}{\tau_U} + \frac{\hat{W}_R(u,h)}{\tau_R} + \frac{\hat{W}_H}{\tau_H}\right] P, \tag{S27}$$

where $\tau_U \ll \tau_R \ll \tau_H$, since we are assuming, as before, that U is the fastest variable. Here, $\varepsilon = \tau_U/\tau_R$ and $\delta = \tau_R/\tau_H$. Notice that now $\hat{W}_R$ depends also on u. We can solve the system again by resorting to a timescale separation and scaling the time by the slowest timescale, τ_H. We have:

$$\partial_t P = \left[\epsilon^{-1}\delta^{-1}\hat{W}_U(r) + \delta^{-1}\hat{W}_R(u,h) + \hat{W}_H\right] P\,. \tag{S28}$$

We now expand the propagator at first order in ε, $P = P^{(0)} + \varepsilon P^{(1)}$. Then, the order ε^{-1} of the master equation gives, as above, $P^{(0)} = p^{\mathrm{st}}_{U|R}(u|r)\Pi(r,h,t|r_0,h_0,t_0)$. At order ε^0, Eq. (S28) leads to

$$\partial_t \Pi(r,h,t|r_0,h_0,t_0) = \left[\delta^{-1}\sum_u \hat{W}_R(u,h)p^{\mathrm{st}}_{U|R}(u|r) + \hat{W}_H\right]\Pi(r,h,t|r_0,h_0,t_0). \tag{S29}$$

To solve this, we expand the propagator as $\prod = \prod^{(0)} + \delta\prod^{(1)}$ and, at order δ^{-1}, we obtain:

$$\left(\sum_u \hat{W}_R(u,h)p^{\mathrm{st}}_{U|R}(u|r)\right)\Pi^{(0)} = 0 \tag{S30}$$

This is a 2 × 2 effective matrix acting on $\prod^{(0)}$, where the only rate affected by u is $\Gamma^{(C)}_{A\to P}$, which multiplies the active states, i.e., $r = 1$. This equation can be analytically computed and the solution of Eq. (S30) is:

$$\Pi^{(0)} = \rho^{\mathrm{st}}_{R|H}(r|h)f(h,t|h_0t_0) \qquad \rho^{\mathrm{st}}_{R|H}(r=0|h) = \frac{e^{\beta\Delta E}(1+\Theta)}{e^{\beta h}+1+e^{\beta\Delta E}(1+\Theta)} \tag{S31}$$

with $\log(\Theta) = e^{-\beta(V-c)}(e^{\beta\theta} - 1)$. Clearly, $\rho^{\mathrm{st}}_{R|H}(r|h)$ does not depend on u since we summed over the fast variable. Going on with the computation, at order δ^0, we obtain:

$$\partial_t f(h,t|h_0,t_0) = \hat{W}_H f(h,t|h_0,t_0) \tag{S32}$$

So that the full propagator results to be:

$$P^{(0)}(u,r,h,t|u_0,r_0,h_0,t_0) = p^{\mathrm{st}}_{U|R}(u|r)\rho^{\mathrm{st}}_{R|H}(r|h)P_H(h,t|h_0,t_0) \tag{S33}$$

From this expression, we can find the joint probability distribution, following the same steps as before:

$$p_{U,R,H}(u,r,h,t) = p^{\mathrm{st}}_{U|R}(u|r)\rho^{\mathrm{st}}_{R|H}(r|h)p_H(h,t) \tag{S34}$$

As expected, since U relaxes instantaneously, the feedback is instantaneous as well. As a consequence, the timedependent behavior of the system is solely driven by the external field H, with a fixed amplitude that takes into account the effect of the feedback only on average. This means that there will be no dynamic reduction of activity and, as such, no habituation in this scenario. This was somehow expected, since all variables are faster than the external field and, as a consequence, the feedback cannot be implemented over time. The first conclusion is that the variable implementing the feedback has to evolve together with H.

2. Effective dynamical feedback requires an additional population

We now assume that the feedback is, again, implemented by U, but it acts as a slow variable. Formally, we take $\tau_R \ll \tau_U \approx \tau_H$. Rescaling the time by the slowest timescale, τ_H (works the same for τ_U), we have:

$$\partial_t P = \left[\frac{\tau_H}{\tau_U}\hat{W}_U(r) + \epsilon^{-1}\hat{W}_R(u,h) + \hat{W}_H\right]P \tag{S35}$$

with $\varepsilon = \tau_R/\tau_H$. We now expand the propagator at first order in ε, $P = P^{(0)} + \varepsilon P^{(1)}$. Then, the order ε^{-1} of the master equation is simply $\hat{W}_R P^{(0)} = 0$, whose solution gives $P^{(0)} = p^{\mathrm{st}}_{R|U,H}(r|u,h)\Pi(u,h,t|u_0,h_0,t_0)$. At order ε^0:

$$\partial_t \Pi(u,h,t|u_0,h_0,t_0) = \left[\frac{\tau_H}{\tau_U}\sum_r \hat{W}_U(r)p^{\mathrm{st}}_{R|U,H}(r|u,h) + \hat{W}_H\right]\Pi(u,h,t|u_0,h_0,t_0). \tag{S36}$$

The only dependence on r in $\hat{W}_U(r)$ is through the production rate of U. Indeed, the effective transition matrix governing the birth-and-death process of readout molecules is characterized by:

$$\Gamma^{\mathrm{eff}}_{u\to u+1} = e^{-\beta V}\left(e^{\beta c}p^{\mathrm{st}}_{R|U,H}(r=1|u,h) + p^{\mathrm{st}}_{R|U,H}(r=0|u,h)\right)\Gamma^0_U \tag{S37}$$

This rate depends only on h, but h evolves in time. Therefore, we should scan all possible (infinite) values that h takes and build an infinite dimensional transition matrix. In order to solve the system, imagine that we are looking at the interval $[t_0, t_0 + \Delta t]$. Then, we can employ the following approximation if $\Delta t \ll \tau_H$:

$$\Gamma^{\text{eff}}_{u\to u+1}(h) = \Gamma^{\text{eff}}_{u\to u+1}(h_0) \tag{S38}$$

Using this simplification, we need to solve the following equation:

$$\partial_t \Pi(u, h, t_0 + \Delta t|u_0, h_0, t_0) = \left[\frac{\tau_H}{\tau_U}\hat{W}^{\text{eff}}_U(u, h_0) + \hat{W}_H\right]\Pi(u, h, t_0 + \Delta t|u_0, h_0, t_0). \tag{S39}$$

The explicit solution in the interval $t \in [t_0, t_0 + \Delta t]$ can be found to be:

$$\Pi(u, h, t_0 + \Delta t|u_0, h_0, t_0) = P^{\text{eff}}_U(u, t_0 + \Delta t|u_0, h_0, t_0)P_H(h, t_0 + \Delta t|h_0, t_0) \tag{S40}$$

with P^{eff}_U a propagator. The full propagator at time $t_0 + \Delta t$ is then:

$$p_{U,R,H}(u, r, h, t_0 + \Delta t|u_0, r_0, h_0, t_0) = \sum_{u_0} p^{\text{st}}_{R|U,H}(r|u, h)P^{\text{eff}}_U(u, t_0 + \Delta t|u_0, h_0, t_0)P_H(h, t_0 + \Delta t|h_0, t_0)p_{U,H}(u_0, h_0, t_0) \tag{S41}$$

Integrating over the initial conditions, we finally obtain:

$$p_{U,R,H}(u, r, h, t_0 + \Delta t) = \sum_{u_0} p^{\text{st}}_{R|U,H}(r|u, h)\int dh_0 P^{\text{eff}}_U(u, t_0 + \Delta t|u_0, h_0, t_0)P_H(h, t_0 + \Delta t|h_0, t_0)p_{U,H}(u_0, h_0, t_0) \tag{S42}$$

To numerically integrate this equation, we make two approximations. The first one is that we solve the dynamics in all intervals in which the field does not evolve, where P_H is a delta function peaked at the initial condition. For all time points in which the field changes, this amounts to considering the field at the previous instant, a good approximation as long $\Delta t \ll \tau_H$, particularly when the time dependence of the field is a square wave, as in our case.

The second approximation is to compute the propagator of P_U. As explained in the Methods of the main text, we restrict our computation to the transitions between n nearest neighbors in the U space. In the case of transitions only among next-nearest neighbors, we have the following dynamics:

$$\partial_t P(u|u_0, h) = W^{\text{nn}}P(u|u_0) \tag{S43}$$

with the transition matrix:

$$\begin{aligned} W^{\text{nn}}_{12} &= \hat{W}_{u_0\to u_0-1} = \Gamma^U_0 u_0 & W^{\text{nn}}_{13} &= \hat{W}_{u_0+1\to u_0-1} = 0 \\ W^{\text{nn}}_{21} &= \hat{W}_{u_0-1\to u_0} = \Gamma^{\text{eff}}_{u_0-1\to u_0} & W^{\text{nn}}_{23} &= \hat{W}_{u_0+1\to u_0} = \Gamma^U_0(u_0 + 1) \\ W^{\text{nn}}_{31} &= \hat{W}_{u_0+1\to u_0-1} = 0 & W^{\text{nn}}_{32} &= \hat{W}_{u_0\to u_0+1} = \Gamma^{\text{eff}}_{u_0\to u_0+1} \end{aligned}$$

the diagonal is fixed to satisfy the conservation of normalization, as usual. The solution is:

$$P(u|u_0,h) = p^{\rm st}_{U|H} + \sum_\nu w_\nu a^{(\nu)} e^{\lambda_\nu \Delta t} \tag{S44}$$

where w_ν and λ_ν are respectively eigenvectors and eigenvalues of the transition matrix $W^{\rm nn}$. The coefficients $a^{(\nu)}$ have to be evaluated according to the condition at time t^0:

$$P_{U|H}(u|u_0,h) = p^{\rm st}_{U|H} + \sum_\nu w_\nu a^{(\nu)} = \delta_{u,u_0} \tag{S45}$$

where $\delta_{u,\,u0}$ is the Kroenecker's delta. To evaluate the information content of this model, we also need:

$$\begin{aligned} p_U(u, t_0+\Delta t) &= \sum_{u_0} p_U(u_0,t_0) \int dh P^{\rm eff}_U(u, t_0+\Delta t|u_0,h,t_0) p_H(h, t_0+\Delta t) \\ p_{U|H}(u, t_0+\Delta t|h) &= \sum_{u_0} P^{\rm eff}_U(u, t_0+\Delta t|u_0,h,t_0) p_U(u_0,t_0)\ . \end{aligned} \tag{S46}$$

In **Figure S6** we show that, in this model, U does not display habituation. Rather, it increases upon repeated stimuli, acting as the storage in the main text. On the other hand, the probability of the receptor being active does habituate. This suggests that habituation can only occur in fast variables modulated by slow variables.

It is straightforward to intuitively understand why a direct feedback from U, with this population undergoing a slow dynamics, cannot lead to habituation. Indeed, at a fixed distribution of the external signal, the stationary solution of $\langle U\rangle$ already takes into account the effect of the negative feedback. Hence, if the system starts with a very low readout population (no signal), the dynamics induced by a switching signal can only bring $\langle U\rangle$ to its steady state with intervals in which the population will grow and intervals in which it decreases. Naively speaking, the dynamics of $\langle U\rangle$ becomes similar to the one of the storage in the complete model, since it is actually playing the same role of storing information in this simplified context.

S5. Robustness of Optimality

1. Effects of the external signal strength and thermal noise level

In the main text, for analytical ease, we take the environment to be an exponentially distributed signal,

$$p_H(h,t) = \lambda(t) e^{-h\lambda(t)} \tag{S47}$$

where λ is its inverse characteristic scale. In particular, we describe the case in which no signal is present by setting λ to be large, so that the typical realizations of H would be too small to activate the receptors. On the other hand, when λ is small, the values of h appearing in the rates of the model are large enough to activate the receptor and thus allow the system to sense the signal.

In the dynamical case, we take $\lambda(t)$ to be a square wave, so that $\langle H\rangle = 1/\lambda$ alternates between two values $\langle H\rangle_{\rm min}$ and $\langle H\rangle_{\rm max}$. We denote with $T_{\rm on}$ the duration of $\langle H\rangle_{\rm max}$, and with $T_{\rm off}$ the one of $\langle H\rangle_{\rm min}$. In practice, this signal mimics an on-off dynamics, where the stochastic signal is present only when its average is large enough, $\langle H\rangle_{\rm max}$. In the main text, we take $\langle U\rangle_{\rm min} = 0.1$ and $\langle H\rangle_{\rm max} = 10$, with $T_{\rm on} = T_{\rm off} = 100\Delta t$.

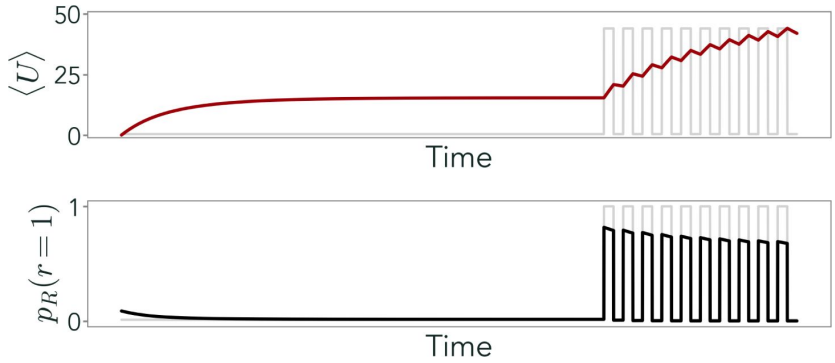


FIG. S6.

Dynamics of a system where *U* evolves on the same timescale of *H*, and implements directly a negative feedback on the receptor. In this model, ⟨*U*⟩ (in red) increases upon repeated stimulation rather than decreasing, responding to changes in ⟨*H*⟩ (in gray) as the storage of the full model. On the other hand, the probability of the receptor being active, *pR*(*r* = 1) (black), shows signs of habituation.

In **Figure S7a**, we study the behavior of the model in the presence of a static exponential signal, with average $\langle H \rangle$. We focus on the case of low σ, so that the production of storage is favored. As $\langle H \rangle$ decreases, $I_{U,H}$ decreases as well. Hence, as expected, information acquired through sensing depends on the strength of the external signal that coincides with the energy input driving receptor activation. However, the system does not display for all parameters an emergent information dynamics, memory, and habituation. In **Figure S7b**, we see that, when the temperature is low but σ is high, the system does not show habituation and $\Delta I_{U,H} = 0$. On the other hand, when thermal noise dominates (**Figure S7c**), even when the external signal is small, the system produces a large readout population due to random thermal activation. As a consequence, these random activation hinders the signal-driven ones, thus the system do not effectively sense the external signal even when present and $I_{U,H}$ is always small. It is important to remind here that, as we see in the main text in **Figure 3b**, at fixed σ and as a function of ς, $I_{U,H}$ is not monotonic. This is due to the fact that low temperatures typically favor sensing and habituation, but they also intrinsically suppress readout production. Thus, at high β, σ needs to be small to effectively store information since thermal noise is negligible. Vice versa, a small σ is detrimental at high temperatures since the system produces storage as a consequence of thermal noise. This complex interplay is captured by the Pareto optimization, which gives us an effective relation between β and σ to maximize storage while minimizing dissipation.

2. 3-dimensional Pareto-like surface

In line with the front derived above, it is possible to maximize both $I_{U,H}$ and ΔI_f, while minimizing δQ_R to study the region of parameters where habituation spontaneously emerges. In this case, the idea is to maximize all the features associated with the capability to process information, while maintaining a minimal dissipation. Since, as expected, $I_{U,H}$ and ΔI_f are not in trade-off (see **Figure S8**), the resulting optimal area will be named Pareto-like surface. In **Figure S8a**, we represent it in the three-dimensional features space. **Figure S8b-d** only represents the pro jection of this area (in gray) into the parameters space, (β, σ). In what follows, we study the robustness of the optimality of our model by showing both the 2-dimensional Pareto front (see main text) and the 3-dimensional Pareto-like surface in the (β, σ) space. In fact, it is immediate to see from what we show below that the two only slightly differ, and considering one or the other does not qualitatively change the results and the conclusions of our work.

3. Static and dynamical optimality

In **Figure S9**, we plot the average readout population, $\langle U \rangle$, the average storage population, $\langle S \rangle$, the mutual information between them, $I_{U,S}$, and the entropy production of the internal processes, $\dot{\Sigma}_{\mathrm{int}}$, as a function of β and σ and in the presence of a static field. The optimal values of β and σ obtained by minimizing the Pareto-like functional

$$\mathcal{L}(\beta,\sigma) = \gamma_1 \frac{I_{U,H}(\beta,\sigma)}{\max(I_{U,H})} + \gamma_2 \frac{\Delta I_f(\beta,\sigma)}{\max(\Delta I_f)} - (1-\gamma_1-\gamma_2)\frac{\delta Q_R(\beta,\sigma)}{\max(\delta Q_R)}, \quad \gamma_1+\gamma_2 = 1,$$

are such that both $\langle U \rangle$ and $\langle S \rangle$ attain intermediate values. We also show the 2-dimensional Pareto front derived in the main text for comparison (dashed black line). Thus, large/small production of readout/storage is detrimental to sensing. Similar considerations can be drawn for the dissipation of internal processes, $\dot{\Sigma}_{\mathrm{int}}$. Interestingly, the dependence between readout and storage, quantified by the mutual information $I_{U,S}$, is not maximized at optimality. This suggests that excessively strong negative feedback impedes information, while promoting dissipation in the receptor, δQ_R, thus being suboptimal.

FIG. S7.

Effects of the external signal strength and thermal noise level on sensing. (a) At fixed and low $\sigma = 0.1$ and constant exponentially distributed signal with mean $\langle H \rangle$. As $\langle H \rangle$ decreases, the system captures less information and it needs to operate at lower temperatures to sense the signal. In particular, as the temperature decreases, $I_{U,H}$ becomes larger. (b) In the dynamical case, outside the optimal surface, at high β and high σ, storage is not produced and thus no negative feedback is present. The system does not display habituation, and $I_{U,H}$ is smaller than inside the optimal surface (gray area). (c) In the opposite regime, at low β and σ, the system is dominated by thermal noise. As a consequence, the average readout $\langle U \rangle$ is high even when the external signal is not present ($\langle H \rangle = \langle H \rangle_{\min} = 0.1$), and it captures only a small amount of information $I_{U,H}$, which is masked by thermal activation. Other simulation parameters for this figure are $\langle U \rangle_A = 150$, $\langle U \rangle_P = 0.5$, $\beta = 2/3$, and $\Gamma_S^0 = \Delta E = g$. For the dynamical case, $T_{on} = T_{off} = 100\Delta t$.

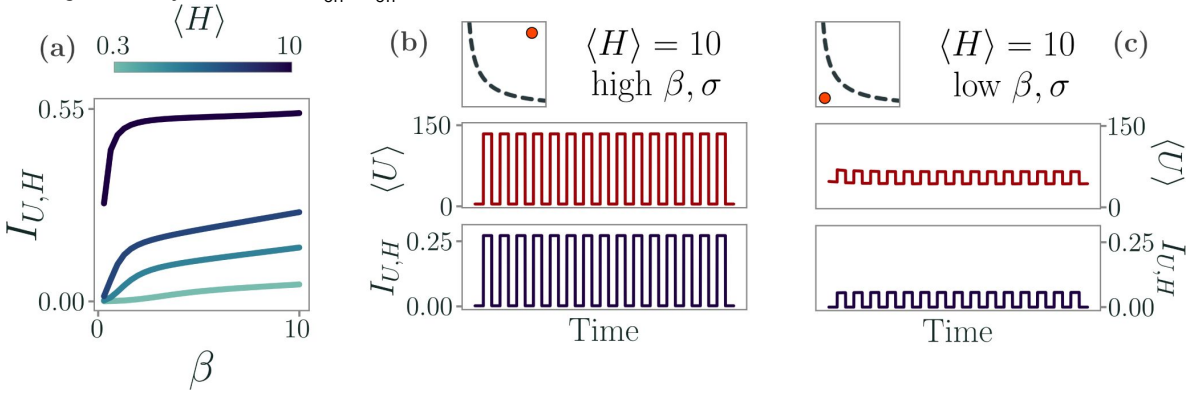
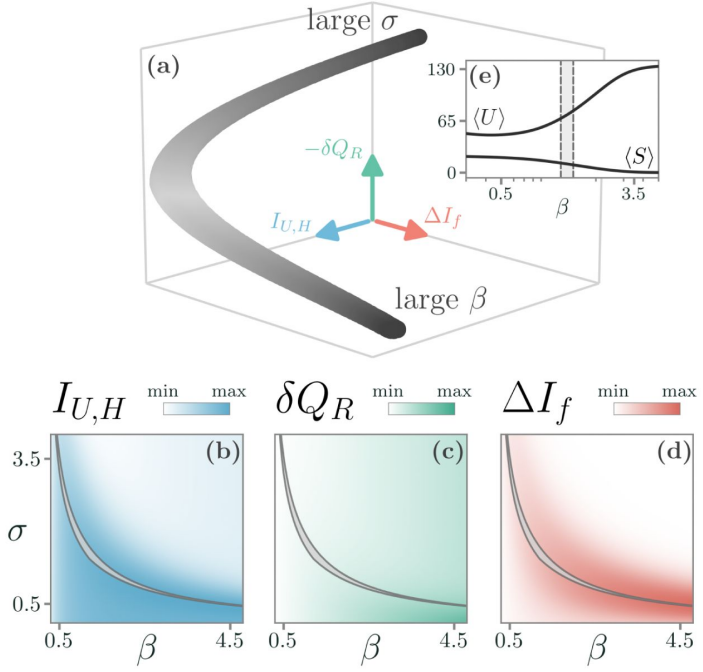


FIG. S8.

Trade-off between energy dissipation and information with a 3*D* optimization including information feedback. (a) The optimal surface in the ($I_{U,H}$, ΔI_f, $-\delta Q_R$) space with a constant external field, is obtained through a Pareto-like optimization. (b-d) The values of σ and β inside the optimal surface (gray surface) maximize both the readout information, $I_{U,H}$, and the information feedback of the storage population, ΔI_f, while minimizing the dissipation of the receptor, δQ_R. (e) At the optimal (β, σ) (gray area, shown at a fixed value of $\sigma \ll 1.5$) the average readout, $\langle U \rangle$, and storage, $\langle S \rangle$, are at intermediate values.



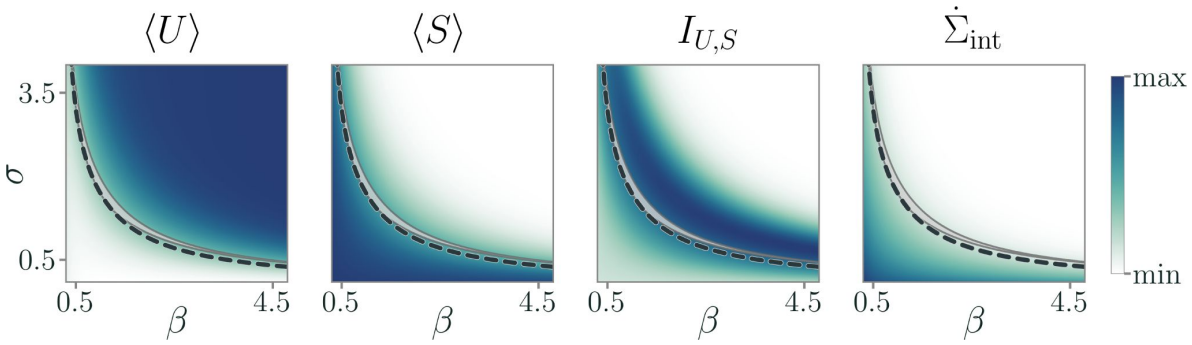


FIG. S9.

Behavior of the average readout population, ⟨*U*⟩, the average storage population, ⟨*S*⟩, the mutual information between them, $I_{U,S}$, and the entropy production of the internal processes, $\dot{\Sigma}_{\mathrm{int}}$, as a function of β and σ and in the presence of a static field. The gray area represents the Pareto-like optimal surface, while the dashed black line indicates the 2-dimensional Pareto front derived in the main text. The signal is exponentially distributed with an inverse characteristic scale $\lambda = 0.1$, so that $\langle H \rangle = 10$. Other simulation parameters are as in **Figure S7**.

In **Figure S10**, we study the dynamical behavior of the model under a repeated external signal, as in **Figure 3f-g-h** in the main text. In particular, given an observable O, we define its change under a repeated signal ΔO as the difference between the maximal response to the signal at large times and the maximal response to the first signal (**Figure S10a**). In **Figure S10b** we see in particular that $\Delta\langle U\rangle$ is maximal in the region where the change in information feedback $\Delta\Delta I_f$ is negative, suggesting that a strong habituation fueled by a large storage concentration is ultimately detrimental for information processing. Furthermore, in this region the entropy produced by internal processes, $\dot{\Sigma}_{\mathrm{int}}$, is maximal.

4. Interplay between information storage and signal duration

In the main text and insofar, we have always considered the case $T_{\mathrm{on}} = T_{\mathrm{off}}$. We now study the effect of the signal duration and the pause length on sensing (**Figure S11**). If the system only receives short signals between long pauses, the slow storage build-up does not reach a high level of concentration. As a consequence, the negative feedback on the receptor is less effective and habituation is suppressed (**Figure S11a**). Therefore, the peak of $\Delta I_{U,H}$ in the (β, σ) plane takes place below the optimal surface, as σ needs to be smaller than in the static case to boost storage production during the brief periods in which the signal is present. On the other hand, in **Figure S11b** we consider the case of a long signal with short pauses. In this scenario, the slow dynamical evolution of the storage can reach large concentrations at larger values of σ, thus moving the optimal dynamical region slightly above the Pareto-like surface. Considering the 2-dimensional Pareto front as a reference, it does not change qualitatively our observations. The case of a short signal is comparable to the durations of the looming stimulations in the experimental setting (see next Section), which can be used to tune the parameters of the model to the peak of information gain.

S6. Experimental Setup

Acquisitions of the zebrafish brain activity were carried out in Elavl3:H2BGCaMP6s larvae at 5 days post fertilization raised at 28°C on a 12 h light/12 h dark cycle according to the approval by the Ethical Committee of the University of Padua (61/2020 dal Maschio). Larvae were embedded in 2 percent agarose gel and their brain activity was recorded using a multiphoton system with a custom 3D volumetric acquisition module. Briefly, the imaging path is based on an 8-kHz galvo-resonant commercial 2P design (Bergamo I Series, Thorlabs, Newton, NJ, United States) coupled to a Ti:Sapphire source (Chameleon Ultra II, Coherent) tuned to 920nm for imaging GCaMP6 signals and modulated by a Pockels cell(Conoptics). The fluorescence collection path includes a 705 nm long-pass main dichroic and a 495nm long-pass dichroic mirror transmitting the fluorescence light toward a GaAsP PMT detector (H7422PA-40, Hamamatsu) equipped with EM525/50 emission filter. Data were acquired at 30 frames per second, using a water dipping Nikon CFI75 LWD 16X W objective covering an effective field of view of about 450 × 900um with a resolution of 512 × 1024 pixels. The volumetric module is based on an electrically tunable lens (Optotune) moving continuously according to a saw-tooth waveform synchronized with the frame acquisition trigger. An entire volume of about 180 – 200um in thickness encompassing 30 planes separated by about 7um is acquired at a rate of 1 volume per second, sufficient to track the relative slow dynamics associated with the fluorescence-based activity reporter GCaMP6s.

As for the visual stimulation, looming stimuli were generated using Stytra and presented monocularly on a 50 × 50mm screen using a DPL4500 projector by Texas Instruments. The dark looming dot was presented 10 times with 150s interval, centered with the fish eye and with a l/v parameter of 8.3 s, reaching at the end of the stimulation a visual angle of 79.4° corresponding to an angular expansion rate of 9.5°/s. The acquired temporal series were first processed using an automatic pipeline, including motion artifact correction, temporal filtering with a rectangular

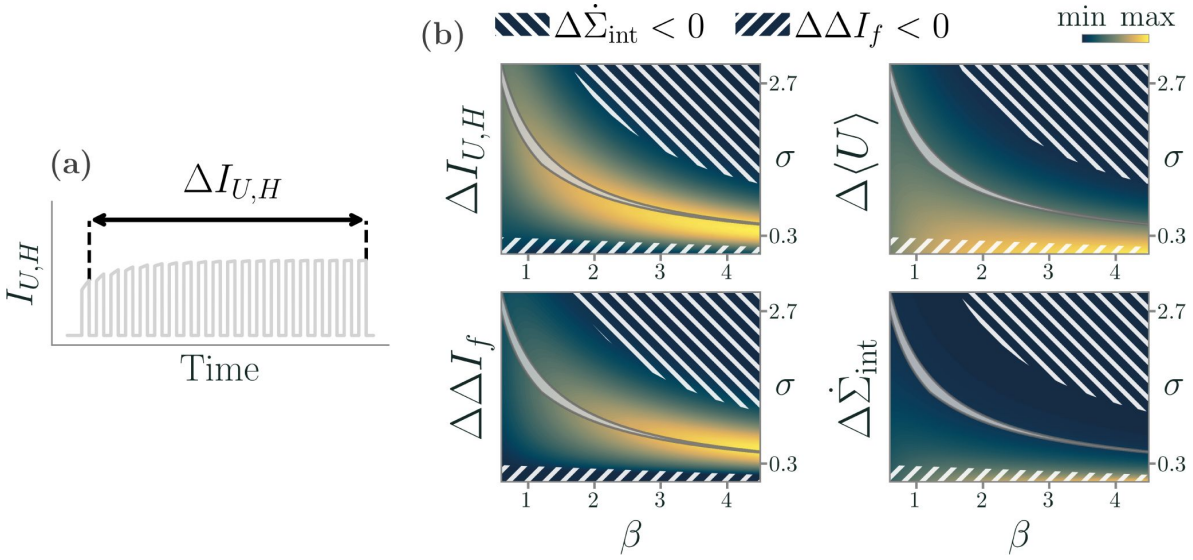


FIG. S10.

Dynamical optimality under a repeated external signal. (a) Schematic definition of how we study the dynamical evolution of relevant observables, by comparing the maximal response to a first signal with the one to a signal at large times. (b) Behavior of the increase in readout information, $\Delta I_{U,H}$, in feedback information, $\Delta\Delta I_f$, in average storage population, $\Delta\langle U\rangle$, and in entropy production, $\Delta\ \dot{\Sigma}_{\rm int}$. The gray area represents the Pareto-like optimal surface in the presence of a static field, while the dashed black line indicates the 2-dimensional Pareto front obtained in the same conditions. Simulations parameters are as in **Figure S7**. In particular, recall the signal is exponentially distributed whose characteristic scale follows a square wave, with $\langle H\rangle_{\rm max} = 10$, $\langle H\rangle_{\rm max} = 0.1$, and $T_{\rm on} = T_{\rm off} = 100\Delta t$.

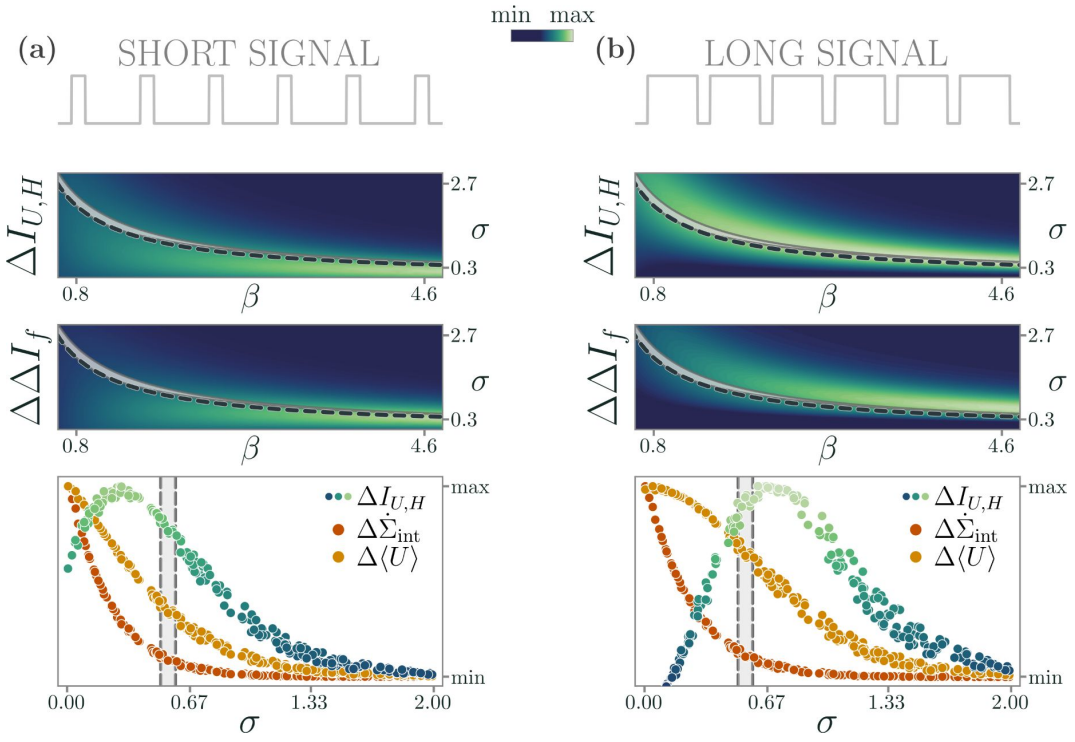


FIG. S11.

Effect of the signal duration on habituation. (a) If the system only receives the signal for a short time ($T_{on} = 50\Delta T < T_{off} = 200\Delta T$) it does not have enough time to reach a high level of storage concentration. As a consequence, both ΔU and $\Delta I_{U,H}$ are smaller, and thus habituation is less effective. (b) If the system receives long signals with brief pauses ($T_{on} = 200\Delta T > T_{off} = 50\Delta T$), instead, the habituation mechanism promotes information storage and thus a reduction in the readout activity. Other simulation parameters are as in **Figure S10**. Gray area and dashed black line represent the 3dimensional and 2dimensional optimal region where habituation emerges respectively.

window 3 second long, and automatic segmentation using Suite2P. Then, the obtained dataset was manually curated to resolve segmentation errors or to integrate cells not detected automatically. We fit the activity profiles of about 52,000 cells with a linear regression model (scikit-learn Python Library) using a set of base functions representing the expected responses to each of the stimulation events, obtained convolving an exponentially decaying kernel of the GCaMP signal lifetime with square waveforms characterized by an amplitude different from zero only during the presentation of the corresponding visual stimulus. The resulting coefficients were divided for the mean squared error of the fit to obtain a set of scores. The cells, whose score fell within the top 5 of the distribution, were considered for the dimensionality reduction analysis.

The resulting fluorescence signals $F^{(i)}$, for $i = 1, \dots, N_{\text{cells}}$, were processed by removing a moving baseline to account for baseline drifting and fast oscillatory noise [74]. Briefly, for each time point t, we selected a window $[t - \tau_2, t]$ and evaluated the minimum smoothed fluorescence,

$$F_0^{(i)} = \min_{u\in[t-\tau_2,t]} \left[\frac{1}{\tau_1} \int_{u-\tau_1/2}^{u+\tau_1/2} F(s)\, ds \right]. \tag{S48}$$

Then, the relative change in fluorescence signal,

$$R^{(i)}(t) = \frac{F^{(i)}(t) - F_0^{(i)}}{F_0^{(i)}} \tag{S49}$$

is smoothed with an exponential moving average. Thus, the neural activity profile for the i-th cell that we use in the main text is given by

$$x^{(i)}(t) = \frac{\int_0^t R(t-\tau) w(\tau) d\tau}{\int_0^t w(\tau) d\tau}, \quad w(t) = \exp\left[-\frac{t}{\tau_0}\right]. \tag{S50}$$

In accordance with the previous literature [74], we set $\tau_0 = 0.2$ s, $\tau_1 = 0.75$ s, and $\tau_2 = 3$ s. The qualitative nature of the low-dimensional activity in the PCA space is not altered by other sensible choices of these parameters.

References

[1] Tkacik G., Bialek W. (2016) **Information processing in living systems** *Annual Review of Condensed Matter Physics* **7**

[2] Azeloglu E. U., Iyengar R. (2015) **Signaling networks: information flow, computation, and decision making** *Cold Spring Harbor perspectives in biology* **7**

[3] Gnesotto F. S., Mura F., Gladrow J., Broedersz C. P. (2018) **Broken detailed balance and non-equilibrium dynamics in living systems: a review** *Reports on Progress in Physics* **81**

[4] Nemenman I. (2012) **Information theory and adaptation** *Quantitative biology: from molecular to cellular systems* **4**

[5] Nakajima T. (2015) **Biologically inspired information theory: Adaptation through construction of external reality models by living systems** *Progress in Biophysics and Molecular Biology* **119**

[6] Whiteley M., Diggle S. P., Greenberg E. P. (2017) **Progress in and promise of bacterial quorum sensing research** *Nature* **551**

[7] Perkins T. J., Swain P. S. (2009) **Strategies for cellular decision-making** *Molecular systems biology* **5**

[8] Koshland D. E., Goldbeter A., Stock J. B. (1982) **Amplification and adaptation in regulatory and sensory systems** *Science* **217**

[9] Tu Y., Shimizu T. S., Berg H. C. (2008) **Modeling the chemotactic response of escherichia coli to time-varying stimuli** *Proceedings of the National Academy of Sciences* **105**

[10] Tu Y. (2008) **The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by maxwell's demons** *Proceedings of the National Academy of Sciences* **105**

[11] Mattingly H., Kamino K., Machta B., Emonet T. (2021) **Escherichia coli chemotaxis is information limited** *Nature Physics* **17**

[12] Cheong R., Rhee A., Wang C. J., Nemenman I., Levchenko A. (2011) **Information transduction capacity of noisy biochemical signaling networks** *science* **334**

[13] Wajant H., Pfizenmaier K., Scheurich P. (2003) **Tumor necrosis factor signaling** *Cell Death & Differentiation* **10**

[14] Marquez-Legorreta E., Constantin L., Piber M., Favre-Bulle I. A., Taylor M. A., Blevins A. S., Giacomotto J., Bassett D. S., Vanwalleghem G. C., Scott E. K. (2022) **Brain-wide visual habituation networks in wild type and fmr1 zebrafish** *Nature Communications* **13**

[15] Fotowat H., Engert F. (2023) **Neural circuits underlying habituation of visually evoked escape behaviors in larval zebrafish** *Elife* **12**

[16] Lan G., Sartori P., Neumann S., Sourjik V., Tu Y. (2012) **The energy-speed-accuracy trade-off in sensory adaptation** *Nature physics* **8**

[17] Menini A. (1999) **Calcium signalling and regulation in olfactory neurons** *Current opinion in neurobiology* **9**

[18] Kohn A. (2007) **Visual adaptation: physiology, mechanisms, and functional benefits** *Journal of neurophysiology* **97**

[19] Lesica N. A., Jin J., Weng C., Yeh C.-I., Butts D. A., Stanley G. B., Alonso J.-M. (2007) **Adaptation to stimulus contrast and correlations during natural visual stimulation** *Neuron* **55**

[20] Benucci A., Saleem A. B., Carandini M. (2013) **Adaptation maintains population homeostasis in primary visual cortex** *Nature neuroscience* **16**

[21] Schneidman E., Berry M. J., Segev R., Bialek W. (2006) **Weak pairwise correlations imply strongly correlated network states in a neural population** *Nature* **440**

[22] Tkacik G., Marre O., Amodei D., Schneidman E., Bialek W., Berry M. J. (2014) **Searching for collective behavior in a large network of sensory neurons** *PLoS computational biology* **10**

[23] Kurtz Z. D., Müller C. L., Miraldi E. R., Littman D. R., Blaser M. J., Bonneau R. A. (2015) **Sparse and compositionally robust inference of microbial ecological networks** *PLoS computational biology* **11**

[24] Tunstrøm K., Katz Y., Ioannou C. C., Huepe C., Lutz M. J., Couzin I. D. (2013) **Collective states, multistability and transitional behavior in schooling fish** *PLoS computational biology* **9**

[25] Nicoletti G., Busiello D. M. (2021) **Mutual information disentangles interactions from changing environments** *Physical Review Letters* **127**

[26] Nicoletti G., Busiello D. M. (2022) **Mutual information in changing environments: non-linear interactions, out-ofequilibrium systems, and continuously-varying diffusivities** *Physical Review E* **106**

[27] De Smet R., Marchal K. (2010) **Advantages and limitations of current network inference methods** *Nature Reviews Microbiology* **8**

[28] Nicoletti G., Maritan A., Busiello D. M. (2022) **Information-driven transitions in projections of underdamped dynamics** *Physical Review E* **106**

[29] Celani A., Shimizu T. S., Vergassola M. (2011) **Molecular and functional aspects of bacterial chemotaxis** *Journal of Statistical Physics* **144**

[30] Kollmann M., Løvdok L., Bartholomé K., Timmer J., Sourjik V. (2005) **Design principles of a bacterial signalling network** *Nature* **438**

[31] de Ronde W. H., Tostevin F., Ten Wolde P. R. (2010) **Effect of feedback on the fidelity of information transmission of time-varying signals** *Physical Review E* **82**

[32] Selimkhanov J., Taylor B., Yao J., Pilko A., Albeck J., Hoffmann A., Tsimring L., Wollman R. (2014) **Accurate information transmission through dynamic biochemical signaling networks** *Science* **346**

[33] Barkai N., Leibler S. (1997) **Robustness in simple biochemical networks** *Nature* **387**

[34] Parrondo J. M., Horowitz J. M., Sagawa T. (2015) **Thermodynamics of information** *Nature physics* **11**

[35] Flatt S., Busiello D. M., Zamuner S., De Los Rios P. (2023) **Abc transporters are billion-year-old maxwell demons** *Communications Physics* **6**

[36] Bilancioni M., Esposito M., Freitas N. (2023) **A chemical reaction network implementation of a maxwell demon** *The Journal of Chemical Physics*

[37] Bennett C. H. (1982) **The thermodynamics of computation — a review** *International Journal of Theoretical Physics* **21**

[38] Sagawa T., Ueda M. (2009) **Minimal energy cost for thermodynamic information processing: measurement and information erasure** *Physical review letters* **102**

[39] Hartich D., Barato A. C., Seifert U. (2015) **Nonequilibrium sensing and its analogy to kinetic proofreading** *New Journal of Physics* **17**

[40] Skoge M., Naqvi S., Meir Y., Wingreen N. S. (2013) **Chemical sensing by nonequilibrium cooperative receptors** *Physical review letters* **110**

[41] Lestas I., Vinnicombe G., Paulsson J. (2010) **Fundamental limits on the suppression of molecular fluctuations** *Nature* **467**

[42] Coultrap S. J., Bayer K. U. (2012) **Camkii regulation in information processing and storage** *Trends in neurosciences* **35**

[43] Frankland P. W., Josselyn S. A. (2016) **In search of the memory molecule** *Nature* **535**

[44] Lisman J., Schulman H., Cline H. (2002) **The molecular basis of camkii function in synaptic and behavioural memory** *Nature Reviews Neuroscience* **3**

[45] Sartori P., Granger L., Lee C. F., Horowitz J. M. (2014) **Thermodynamic costs of information processing in sensory adaptation** *PLoS computational biology* **10**

[46] Barato A. C., Hartich D., Seifert U. (2014) **Efficiency of cellular information processing** *New Journal of Physics* **16**

[47] Ouldridge T. E., Govern C. C., ten Wolde P. R. (2017) **Thermodynamics of computational copying in biochemical systems** *Physical Review X* **7**

[48] Penocchio E., Avanzini F., Esposito M. (2022) **Information thermodynamics for deterministic chemical reaction networks** *The Journal of Chemical Physics*

[49] Rahi S. J., Larsch J., Pecani K., Katsov A. Y., Mansouri N., Tsaneva-Atanasova K., Sontag E. D., Cross F. R. (2017) **Oscillatory stimuli differentiate adapting circuit topologies** *Nature methods* **14**

[50] Tadres D., Wong P. H., To T., Moehlis J., Louis M. (2022) **Depolarization block in olfactory sensory neurons expands the dimensionality of odor encoding** *Science Advances* **8**

[51] Jalaal M., Schramma N., Dode A., de Maleprade H., Raufaste C., Goldstein R. E. (2020) **Stress-induced dinoflagellate bioluminescence at the single cell level** *Physical Review Letters* **125**

[52] Malmierca M. S., Sanchez-Vives M. V., Escera C., Bendixen A. (2014) **Neuronal adaptation, novelty detection and regularity encoding in audition** *Frontiers in Systems Neuroscience* **8** https://doi.org/10.3389/fnsys.2014.00111

[53] Shew W. L., Clawson W. P., Pobst J., Karimipanah Y., Wright N. C., Wessel R. (2015) **Adaptation to sensory input tunes visual cortex to criticality** *Nature Physics* **11**

[54] Lamire L.-A., Haesemeyer M., Engert F., Granato M., Randlett O. (2022) **Inhibition drives habituation of a larval zebrafish visual response** *bioRxiv*

[55] Benda J. (2021) **Neural adaptation** *Current Biology* **31**

[56] Bueti D., Bahrami B., Walsh V., Rees G. (2010) **Encoding of temporal probabilities in the human brain** *Journal of Neuroscience* **30**

[57] Sederberg A. J., MacLean J. N., Palmer S. E. (2018) **Learning to make external sensory stimulus predictions using internal correlations in populations of neurons** *Proceedings of the National Academy of Sciences* **115**

[58] Palmer S. E., Marre O., Berry M. J., Bialek W. (2015) **Predictive information in a sensory population** *Proceedings of the National Academy of Sciences* **112**

[59] Ma W., Trusina A., El-Samad H., Lim W. A., Tang C. (2009) **Defining network topologies that can achieve biochemical adaptation** *Cell* **138**

[60] De Los Rios P., Barducci A. (2014) **Hsp70 chaperones are non-equilibrium machines that achieve ultra-affinity by energy consumption** *Elife* **3**

[61] Astumian R. D. (2019) **Kinetic asymmetry allows macromolecular catalysts to drive an information ratchet** *Nature communications* **10**

[62] Yan J., Hilfinger A., Vinnicombe G., Paulsson J., et al. (2019) **Kinetic uncertainty relations for the control of stochastic reaction networks** *Physical review letters* **123**

[63] Hilfinger A., Norman T. M., Vinnicombe G., Paulsson J. (2016) **Constraints on fluctuations in sparsely characterized biological systems** *Physical review letters* **116**

[64] Nicoletti G., Busiello D. M. (2024) **Information propagation in multilayer systems with higher-order interactions across timescales** *Physical Review X* **14**

[65] Ngampruetikorn V., Schwab D. J., Stephens G. J. (2020) **Energy consumption and cooperation for optimal sensing** *Nature communications* **11**

[66] Seoane L. F., Sole R. (2015) **Phase transitions in pareto optimal complex networks** *Physical Review E* **92**

[67] Bruzzone M., Chiarello E., Albanesi M., Miletto Petrazzini M. E., Megighian A., Lodovichi C., Dal Maschio M. (2021) **Whole brain functional recordings at cellular resolution in zebrafish larvae with 3d scanning multiphoton microscopy** *Scientific reports* **11**

[68] Abbott L. F., Nelson S. B. (2000) **Synaptic plasticity: taming the beast** *Nature neuroscience* **3**

[69] Martin S. J., Grimwood P. D., Morris R. G. (2000) **Synaptic plasticity and memory: an evaluation of the hypothesis** *Annual review of neuroscience* **23**

[70] Hidalgo J., Grilli J., Suweis S., Munoz M. A., Banavar J. R., Maritan A. (2014) **Information-based fitness and the emergence of criticality in living systems** *Proceedings of the National Academy of Sciences* **111**

[71] Busiello D. M., Gupta D., Maritan A. (2020) **Coarsegrained entropy production with multiple reservoirs: Unraveling the role of time scales and detailed balance in biology-inspired systems** *Physical Review Research* **2**

[72] Bo S., Celani A. (2017) **Multiple-scale stochastic processes: decimation, averaging and beyond** *Physics reports* **670**

[73] Schnakenberg J. (1976) **Network theory of microscopic and macroscopic behavior of master equation systems** *Reviews of Modern physics* **48**

[74] Jia H., Rochefort N. L., Chen X., Konnerth A. (2011) **In vivo two-photon imaging of sensory-evoked dendritic calcium signals in cortical neurons** *Nature protocols* **6**

Editors

Reviewing Editor
Arvind Murugan
University of Chicago, Chicago, United States of America

Senior Editor
Aleksandra Walczak
École Normale Supérieure - PSL, Paris, France

Reviewer #1 (Public review):

Summary:

The manuscript by Nicoletti et al. presents a minimal model of habituation, a basic form of non-associative learning, addressing both from dynamical and information theory aspects of how habituation can be realized. The authors identify that negative feedback provided with a slow storage mechanism is sufficient to explain habituation.

Strengths:

The authors combine the identification of the dynamical mechanism with information-theoretic measures to determine the onset of habituation and provide a description of how the system can gain maximum information about the environment.

Weaknesses:

I have several main concerns/questions about the proposed model for habituation and its plausibility. In general, habituation does not only refer to a decrease in the responsiveness upon repeated stimulation but as Thompson and Spencer discussed in Psych. Rev. 73, 16-43 (1966), there are 10 main characteristics of habituation, including (i) spontaneous recovery when the stimulus is withheld after response decrement; dependence on the frequency of stimulation such that (ii) more frequent stimulation results in more rapid and/or more pronounced response decrement and more rapid spontaneous recovery; (iii) within a

stimulus modality, the less intense the stimulus, the more rapid and/or more pronounced the behavioral response decrement; (iv) the effects of repeated stimulation may continue to accumulate even after the response has reached an asymptotic level (which may or may not be zero, or no response). This effect of stimulation beyond asymptotic levels can alter subsequent behavior, for example, by delaying the onset of spontaneous recovery.

These are only a subset of the conditions that have been experimentally observed and therefore a mechanistic model of habituation, in my understanding, should capture the majority of these features and/or discuss the absence of such features from the proposed model.

Furthermore, the habituated response in steady-state is approximately 20% less than the initial response, which seems to be achieved already after 3-4 pulses, the subsequent change in response amplitude seems to be negligible, although the authors however state "after a large number of inputs, the system reaches a time-periodic steady-state". How do the authors justify these minimal decreases in the response amplitude? Does this come from the model parametrization and is there a parameter range where more pronounced habituation responses can be observed?

The same is true for the information content (Figure 2f) - already at the first pulse, IU, H ~ 0.7 and only negligibly increases afterwards. In my understanding, during learning, the mutual information between the input and the internal state increases over time and the system extracts from these predictions about its responses. In the model presented by the authors, it seems the system already carries information about the environment which hardly changes with repeated stimulus presentation. The complexity of the signal is also limited, and it is very hard to clarify from the presented results, whether the proposed model can actually explain basic features of habituation, as mentioned above.
Additionally, there have been two recent models on habituation and I strongly suggest that the authors discuss their work in relation to recent works (bioRxiv 2024.08.04.606534; arXiv:2407.18204).

https://doi.org/10.7554/eLife.99767.1.sa2

Reviewer #2 (Public review):

In this study, the authors aim to investigate habituation, the phenomenon of increasing reduction in activity following repeated stimuli, in the context of its information-theoretic advantage. To this end, they consider a highly simplified three-species reaction network where habituation is encoded by a slow memory variable that suppresses the receptor and therefore the readout activity. Using analytical and numerical methods, they show that in their model the information gain, the difference between the mutual information between the signal and readout after and before habituation, is maximal for intermediate habituation strength. Furthermore, they demonstrate that the Pareto front corresponds to an optimization strategy that maximizes the mutual information between signal and readout in the steady state, minimizes some form of dissipation, and also exhibits similar intermediate habituation strength. Finally, they briefly compare predictions of their model to whole-brain recordings of zebrafish larvae under visual stimulation.

The author's simplified model might serve as a solid starting point for understanding habituation in different biological contexts as the model is simple enough to allow for some analytic understanding but at the same time exhibits all basic properties of habituation in sensory systems. Furthermore, the author's finding of maximal information gain for intermediate habituation strength via an optimization principle is, in general, interesting. However, the following points remain unclear or are weakly explained:

(1) Is it unclear what the meaning of the finding of maximal information gain for intermediate habituation strength is for biological systems? Why is information gain as defined in the paper a relevant quantity for an organism/cell? For instance, why is a system with low mutual information after the first stimulus and intermediate mutual information after habituation better than one with consistently intermediate mutual information? Or, in other words, couldn't the system try to maximize the mutual information acquired over the whole time series, e.g., the time series mutual information between the stimulus and readout?

(2) The model is very similar to (or a simplification of previous models) for adaptation in living systems, e.g., for adaptation in chemotaxis via activity-dependent methylation and demethylation. This should be made clearer.

(3) It remains unclear why this optimization principle is the most relevant one. While it makes sense to maximize the mutual information between stimulus and readout, there are various choices for what kind of dissipation is minimized. Why was \delta Q_R chosen and not, for instance, \dot{\Sigma}_int or the sum of both? How would the results change in that case? And how different are the results if the mutual information is not calculated for the strong stimulation input statistics but for the background one?

(4) The comparison to the experimental data is not too strong of an argument in favor of the model. Is the agreement between the model and the experimental data surprising? What other behavior in the PCA space could one have expected in the data? Shouldn't the 1st PC mostly reflect the "features", by construction, and other variability should be due to progressively reduced activity levels?

https://doi.org/10.7554/eLife.99767.1.sa1

Reviewer #3 (Public review):

The authors use a generic model framework to study the emergence of habituation and its functional role from information-theoretic and energetic perspectives. Their model features a receptor, readout molecules, and a storage unit, and as such, can be applied to a wide range of biological systems. Through theoretical studies, the authors find that habituation (reduction in average activity) upon exposure to repeated stimuli should occur at intermediate degrees to achieve maximal information gain. Parameter regimes that enable these properties also result in low dissipation, suggesting that intermediate habituation is advantageous both energetically and for the purpose of retaining information about the environment.

A major strength of the work is the generality of the studied model. The presence of three units (receptor, readout, storage) operating at different time scales and executing negative feedback can be found in many domains of biology, with representative examples well discussed by the authors (e.g. Figure 1b). A key takeaway demonstrated by the authors that has wide relevance is that large information gain and large habituation cannot be attained simultaneously. When energetic considerations are accounted for, large information gain and intermediate habituation appear to be a favorable combination.

While the generic approach of coarse-graining most biological detail is appealing and the results are of broad relevance, some aspects of the conducted studies, the problem setup, and the writing lack clarity and should be addressed:

(1) The abstract can be further sharpened. Specifically, the "functional role" mentioned at the end can be made more explicit, as it was done in the second-to-last paragraph of the Introduction section ("its functional advantages in terms of information gain and energy

dissipation"). In addition, the abstract mentions the testing against experimental measurements of neural responses but does not specify the main takeaways. I suggest the authors briefly describe the main conclusions of their experimental study in the abstract.

(2) Several clarifications are needed on the treatment of energy dissipation.
- When substituting the rates in Eq. (1) into the definition of δQ_R above Eq. (10), "σ" does not appear on the right-hand side. Does this mean that one of the rates in the lower pathway must include σ in its definition? Please clarify.
- I understand that the production of storage molecules has an associated cost σ and hence contributes to dissipation. The dependence of receptor dissipation on , however, is not fully clear. If the environment were static and the memory block was absent, the term with would still contribute to dissipation. What would be the nature of this dissipation?
- Similarly, in Eq. (9) the authors use the ratio of the rates Γ_{s → s+1} and Γ_{s+1 → s} in their expression for internal dissipation. The first-rate corresponds to the synthesis reaction of memory molecules, while the second corresponds to a degradation reaction. Since the second reaction is not the microscopic reverse of the first, what would be the physical interpretation of the log of their ratio? Since the authors already use σ as the energy cost per storage unit, why not use σ times the rate of producing S as a metric for the dissipation rate?

(3) Impact of the pre-stimulus state. The plots in Figure 2 suggest that the environment was static before the application of repeated stimuli. Can the authors comment on the impact of the pre-stimulus state on the degree of habituation and its optimality properties? Specifically, would the conclusions stay the same if the prior environment had stochastic but aperiodic dynamics?

(4) Clarification about the memory requirement for habituation. Figure 4 and the associated section argue for the essential role that the storage mechanism plays in habituation. Indeed, Figure 4a shows that the degree of habituation decreases with decreasing memory. The graph also shows that in the limit of vanishingly small Δ⟨S⟩, the system can still exhibit a finite degree of habituation. Can the authors explain this limiting behavior; specifically, why does habituation not vanish in the limit Δ⟨S⟩ -> 0?

https://doi.org/10.7554/eLife.99767.1.sa0

Author response:

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The manuscript by Nicoletti et al. presents a minimal model of habituation, a basic form of non-associative learning, addressing both from dynamical and information theory aspects of how habituation can be realized. The authors identify that negative feedback provided with a slow storage mechanism is sufficient to explain habituation.*
>
> *Strengths:*
>
> *The authors combine the identification of the dynamical mechanism with information-theoretic measures to determine the onset of habituation and provide a description of how the system can gain maximum information about the environment.*

We thank the reviewer for highlighting the strength of our work.

> *Weaknesses:*

> *I have several main concerns/questions about the proposed model for habituation and its plausibility. In general, habituation does not only refer to a decrease in the responsiveness upon repeated stimulation but as Thompson and Spencer discussed in Psych. Rev. 73, 16-43 (1966), there are 10 main characteristics of habituation, including (i) spontaneous recovery when the stimulus is withheld after response decrement; dependence on the frequency of stimulation such that (ii) more frequent stimulation results in more rapid and/or more pronounced response decrement and more rapid spontaneous recovery; (iii) within a stimulus modality, the less intense the stimulus, the more rapid and/or more pronounced the behavioral response decrement; (iv) the effects of repeated stimulation may continue to accumulate even after the response has reached an asymptotic level (which may or may not be zero, or no response). This effect of stimulation beyond asymptotic levels can alter subsequent behavior, for example, by delaying the onset of spontaneous recovery.*
>
> *These are only a subset of the conditions that have been experimentally observed and therefore a mechanistic model of habituation, in my understanding, should capture the majority of these features and/or discuss the absence of such features from the proposed model.*

We are really grateful to the reviewer for pointing out these aspects of habituation that we overlooked in the previous version of our manuscript. Indeed, our model is able to capture most of these 10 observed behaviors, specifically: 1) habituation; 2) spontaneous recovery; 3) potentiation of habituation; 4) frequency sensitivity; and 5) intensity sensitivity. Here, we are following the same terminology employed in bioRxiv 2024.08.04.606534, the paper highlighted by the referee. Regarding the hallmark 6) subliminal accumulation, we also believe that our model can capture it as well, but more analyses are needed to substantiate this claim. We will include the discussion of these points in the revised version.

Notably, in line with the discussion in bioRxiv 2024.08.04.606534, we also think that feature 10) long-term habituation, is ambiguous and its appearance might be simply related to the other features discussed above. In the revised version, we will detail our take on this aspect in relation to the presented model.

All other hallmarks require the presence of multiple stimuli and, as a consequence, they cannot be observed within our model, but are interesting lines of research for future investigations. We believe that this addition will help clarify the validity of the model and the relevance of our result, consequently improving the quality of our manuscript.

> *Furthermore, the habituated response in steady-state is approximately 20% less than the initial response, which seems to be achieved already after 3-4 pulses, the subsequent change in response amplitude seems to be negligible, although the authors however state "after a large number of inputs, the system reaches a time-periodic steady-state". How do the authors justify these minimal decreases in the response amplitude? Does this come from the model parametrization and is there a parameter range where more pronounced habituation responses can be observed?*

The referee is correct, but this is solely a consequence of the specific set of parameters we selected. We made this choice solely for visualization purposes. In the next version, when different emerging behaviors characterizing habituation are discussed, we will also present a set of parameters for which habituation can be better appreciated, justifying our new choice.

We stated that the time-periodic steady-state is reached “after a large number of stimuli” from a mathematical perspective. However, by using a habituation threshold, as defined in bioRxiv 2024.08.04.606534 for example, we can say that the system is habituated after a few stimuli for the set of parameters selected in the first version of the manuscript. We will also

discuss this aspect in the Supplemental Material of the revised version, as it will also be important to appreciate the hallmarks of habituation listed above.

> *The same is true for the information content (Figure 2f) - already at the first pulse, IU, H ~ 0.7 and only negligibly increases afterwards. In my understanding, during learning, the mutual information between the input and the internal state increases over time and the system extracts from these predictions about its responses. In the model presented by the authors, it seems the system already carries information about the environment which hardly changes with repeated stimulus presentation. The complexity of the signal is also limited, and it is very hard to clarify from the presented results, whether the proposed model can actually explain basic features of habituation, as mentioned above.*

The point about information is more subtle. We can definitely choose a set of parameters for which the information gain is higher and we will show it in the Supplemental Material of the revised version. However, as the reviewer correctly points out, it is difficult to give an interpretation of the specific value of I_U,H for such a minimal model.

We also remark that, since the readout population and the receptor both undergo a fast dynamics (with appropriate timescales as discussed in the text), we are not observing the transient gain of information associated with the first stimulus and, as such, the mutual information presents a discontinuous behavior resembling the dynamics of the readout.

> *Additionally, there have been two recent models on habituation and I strongly suggest that the authors discuss their work in relation to recent works (bioRxiv 2024.08.04.606534; arXiv:2407.18204).*

We thank the reviewer for pointing out these relevant references. We will discuss analogies and differences in the revised version of the main text. The main difference is the fact that information-theoretic aspects of habituation are not discussed in the presented references, while the idea of this work is to elucidate exactly the interplay between information gain and habituation dynamics.

> ***Reviewer #2 (Public review):***
>
> *In this study, the authors aim to investigate habituation, the phenomenon of increasing reduction in activity following repeated stimuli, in the context of its information-theoretic advantage. To this end, they consider a highly simplified three-species reaction network where habituation is encoded by a slow memory variable that suppresses the receptor and therefore the readout activity. Using analytical and numerical methods, they show that in their model the information gain, the difference between the mutual information between the signal and readout after and before habituation, is maximal for intermediate habituation strength. Furthermore, they demonstrate that the Pareto front corresponds to an optimization strategy that maximizes the mutual information between signal and readout in the steady state, minimizes some form of dissipation, and also exhibits similar intermediate habituation strength. Finally, they briefly compare predictions of their model to whole-brain recordings of zebrafish larvae under visual stimulation.*
>
> *The author's simplified model might serve as a solid starting point for understanding habituation in different biological contexts as the model is simple enough to allow for some analytic understanding but at the same time exhibits all basic properties of habituation in sensory systems. Furthermore, the author's finding of maximal information gain for intermediate habituation strength via an optimization principle is, in general, interesting. However, the following points remain unclear or are weakly explained:*

We thank the reviewer for deeming our work interesting and for considering it a solid starting point for understanding habituation in biological systems.

> *(1) Is it unclear what the meaning of the finding of maximal information gain for intermediate habituation strength is for biological systems? Why is information gain as defined in the paper a relevant quantity for an organism/cell? For instance, why is a system with low mutual information after the first stimulus and intermediate mutual information after habituation better than one with consistently intermediate mutual information? Or, in other words, couldn't the system try to maximize the mutual information acquired over the whole time series, e.g., the time series mutual information between the stimulus and readout?*

This is an important and delicate aspect to discuss. We considered the mutual information with a prolonged stimulation when building the Pareto front, by maximizing this quantity while minimizing the dissipation. The observation that the Pareto front lies in the vicinity of the maximum of the information gain hints at the fact that reducing the information gain by increasing the mutual information at each stimulation will require more energy. However, we did not thoroughly explore this aspect by considering all sources of dissipation and the fact that habituation is, anyway, a dynamical phenomenon. In the revised version, we will clarify this point, extending our analyses.

We would like to add that, from a naive perspective, while the first stimulation will necessarily trigger a certain mutual information, multiple observations of the same stimulus have to reflect into accumulated infor

mation that consequently drives the onset of observed dynamical behaviors, such as habituation.

> *(2) The model is very similar to (or a simplification of previous models) for adaptation in living systems, e.g., for adaptation in chemotaxis via activity-dependent methylation and demethylation. This should be made clearer.*

We apologize for having missed this point. Our choice has been motivated by the fact that we wanted to avoid any confusion between the usual definition of (perfect) adaptation and habituation. At any rate, we will add this clarification in the revised version.

> *(3) It remains unclear why this optimization principle is the most relevant one. While it makes sense to maximize the mutual information between stimulus and readout, there are various choices for what kind of dissipation is minimized. Why was \delta Q_R chosen and not, for instance, \dot{\Sigma}_int or the sum of both? How would the results change in that case? And how different are the results if the mutual information is not calculated for the strong stimulation input statistics but for the background one?*

We thank the referee for giving us the opportunity to deepen this aspect of the manuscript. We decided to minimize \delta Q_R since this dissipation is unavoidable. In fact, considering the existence of two different pathways implementing sensing and feedback, the presence of any input will result in a dissipation produced by the receptor. This energy consumption is reflected in \delta Q_R. Conversely, the dissipation associated with the storage is always zero in the limit of a fast memory. However, we know that such a limit is pathological and leads to no habituation. As a consequence, in the revised version we will discuss other choices for our optimization approach, along with their potentialities and limitations.

The dependence of the Pareto front on the stimulus strength is shown in the Supplemental Material, but not in relation to habituation and information gain. We will strengthen this part

in the revised version of the manuscript, elaborating more on the connection between optimality, information gain, and dynamical behavior.

> *(4) The comparison to the experimental data is not too strong of an argument in favor of the model. Is the agreement between the model and the experimental data surprising? What other behavior in the PCA space could one have expected in the data? Shouldn't the 1st PC mostly reflect the "features", by construction, and other variability should be due to progressively reduced activity levels?*

The agreement between data and model is not surprising - we agree on this - since the data exhibit habituation. However, the fact that, without any explicit biological details, our minimal model is able to capture the features of a complex neural system just by looking at the PCs is non-trivial. The 1st PC only reflects the feature that captures most of the variance of the data and, as such, it is difficult to have a-priori expectations on what it should represent. Depending on the behavior of higher-order PCs, we may include them in the revised version if any interesting results arise.

> ***Reviewer #3 (Public review):***
>
> *The authors use a generic model framework to study the emergence of habituation and its functional role from information-theoretic and energetic perspectives. Their model features a receptor, readout molecules, and a storage unit, and as such, can be applied to a wide range of biological systems. Through theoretical studies, the authors find that habituation (reduction in average activity) upon exposure to repeated stimuli should occur at intermediate degrees to achieve maximal information gain. Parameter regimes that enable these properties also result in low dissipation, suggesting that intermediate habituation is advantageous both energetically and for the purpose of retaining information about the environment.*
>
> *A major strength of the work is the generality of the studied model. The presence of three units (receptor, readout, storage) operating at different time scales and executing negative feedback can be found in many domains of biology, with representative examples well discussed by the authors (e.g. Figure 1b). A key takeaway demonstrated by the authors that has wide relevance is that large information gain and large habituation cannot be attained simultaneously. When energetic considerations are accounted for, large information gain and intermediate habituation appear to be a favorable combination.*

We thank the referee for this positive assessment of our work and its generality.

> *While the generic approach of coarse-graining most biological detail is appealing and the results are of broad relevance, some aspects of the conducted studies, the problem setup, and the writing lack clarity and should be addressed:*
>
> *(1) The abstract can be further sharpened. Specifically, the "functional role" mentioned at the end can be made more explicit, as it was done in the second-to-last paragraph of the Introduction section ("its functional advantages in terms of information gain and energy dissipation"). In addition, the abstract mentions the testing against experimental measurements of neural responses but does not specify the main takeaways. I suggest the authors briefly describe the main conclusions of their experimental study in the abstract.*

We thank the referee for this suggestion. The revised version will present a modified abstract in line with the reviewer's proposal.

> *- When substituting the rates in Eq. (1) into the definition of δQ_R above Eq. (10), "σ" does not appear on the right-hand side. Does this mean that one of the rates in the lower pathway must include σ in its definition? Please clarify.*

We apologize to the referee for this typo. Indeed, \sigma sets the energy scale of the feedback and, as such, it appears in the energetic driving given by the feedback on the receptor, i.e., together with \kappa in Eq. (1). We will fix this issue in the revised version. Moreover, we will check the entire manuscript to be sure that all formulas are consistent.

> *- I understand that the production of storage molecules has an associated cost σ and hence contributes to dissipation. The dependence of receptor dissipation on , however, is not fully clear. If the environment were static and the memory block was absent, the term with would still contribute to dissipation. What would be the nature of this dissipation?*

In the spirit of building a paradigmatic minimal model with a thermodynamic meaning, we considered H to act as an external thermodynamic driving. Since this driving acts on a different pathway with respect to the one affected by the storage, the receptor is driven out of equilibrium by its presence. By eliminating the memory block, we would also be necessarily eliminating the presence of the pathway associated with the storage effect ("internal pathway" in the manuscript). In this case, the receptor is a 2-state, 1-pathway system and, as such, it always satisfies an effective detailed balance. As a consequence, the definition of \delta Q_R reported in the manuscript does not hold anymore and the receptor does not exhibit any dissipation. Our choice to model two different pathways has been biologically motivated. We will make this crucial aspect clearer in the revised manuscript.

> *- Similarly, in Eq. (9) the authors use the ratio of the rates Γ_{s → s+1} and Γ_{s+1 → s} in their expression for internal dissipation. The first-rate corresponds to the synthesis reaction of memory molecules, while the second corresponds to a degradation reaction. Since the second reaction is not the microscopic reverse of the first, what would be the physical interpretation of the log of their ratio? Since the authors already use σ as the energy cost per storage unit, why not use σ times the rate of producing S as a metric for the dissipation rate?*

In the current version of the manuscript, we employed the scheme of a controlled birth and death process to model the coupled process of readout and storage production. Since we are not dealing with a detailed biochemical underlying network, we used this coarse-grained description to capture the main features of the dynamics. In this sense, the considered reactions produce and destroy a molecule from a certain pool even if they are controlled in different ways by the readout. However, we completely agree with the point of view of the referee and will analyze our results following their suggestion.

> *(3) Impact of the pre-stimulus state. The plots in Figure 2 suggest that the environment was static before the application of repeated stimuli. Can the authors comment on the impact of the pre-stimulus state on the degree of habituation and its optimality properties? Specifically, would the conclusions stay the same if the prior environment had stochastic but aperiodic dynamics?*

The initial stimulus is indeed stochastic with an average constant in time. Model response depends on the pre-stimulus level, since it also sets the stationary storage concentration before the first "strong" stimulation arrives. This dependence is not crucial for our result but deserves proper discussion, as the referee correctly pointed out. We will clarify this point in the revised version of this study.

> *(4) Clarification about the memory requirement for habituation. Figure 4 and the associated section argue for the essential role that the storage mechanism plays in habituation. Indeed, Figure 4a shows that the degree of habituation decreases with decreasing memory. The graph also shows that in the limit of vanishingly small Δ⟨S⟩, the system can still exhibit a finite degree of habituation. Can the authors explain this limiting behavior; specifically, why does habituation not vanish in the limit Δ⟨S⟩ -> 0?*

We apologize for the lack of clarity here. Actually, Δ⟨S⟩ is not strictly zero, but equal to 0.15% at the final point. However, due to rounding this appears as 0% in the plot, and we will fix it in the revised version. Let us note that the fact that Δ⟨S⟩ is small signals a nonlinear dependence of Δ⟨U⟩ from Δ⟨S⟩, but no contradiction. We will clarify this aspect in the revised version.

https://doi.org/10.7554/eLife.99767.1.sa4