Latent information about optimal behaviour forces evolution of noisy signal transduction

Age J. Tjalma, Frank J. Bruggeman

Systems Biology Lab, Amsterdam Institute for Molecules, Medicines and Systems, VU University, De Boelelaan 1085, 1081 HV Amsterdam, Netherlands.

Abstract

An open problem in biology is to understand when particular phenotypic adaptation strategies of microorganisms are selected during evolution. They range from random, bet-hedging strategies to deterministic, responsive strategies, relying on signalling circuits. We present an evolutionary model that integrates basic statistical physics of molecular circuits with fitness maximisation and information theory. Besides illustrating when bet-hedging strategies are more evolutionarily successful than responsive strategies, it gives new explanations for several puzzling observations on responsive strategies. For instance, the accuracy with which outputs of signalling networks of single cells track external signals can be remarkably low: cells often distinguish only between 2 to 4 concentration ranges, corresponding to 1 or 2 bits of mutual information between the signal and response variable. Why did evolution lead to such low-fidelity signalling systems? Our theory offers an explanation by taking a novel perspective. It considers the fitness benefit of all signals, including those that are not sensed. We introduce a new concept, 'latent information', which captures the mutual information between all non-sensed signals and the optimal response. The theory predicts that it is often evolutionarily optimal to transduce sensed signals noisily, due to latent information. It indicates that fitness can indeed be maximal when the optimal mutual information extracted from sensed signals is not maximal, but rather has a low value of about 1 or 2 bits - even at moderate values of the latent information - in agreement with experimental findings. Cells likely do not sense all signals because of the fitness cost of expressing many idle signalling systems, which consume limited biosynthetic resources otherwise available for growth. Signals should only be sensed at maximal precision when they contain all information about the optimal response. This work contributes to

a better understanding of the fitness contributions of phenotypic adaptation strategies of microorganisms.

Keywords: Cellular decision making, Phenotypic adaptation strategies, Bet-hedging, Cellular signalling, Information Theory, Mutual information, Biochemical noise

---

1. Introduction

Evolutionary theories about the fitness of phenotypic strategies [1, 2, 3, 4, 5, 6] have made successful predictions of the outcomes of natural selection in fluctuating environments [3, 4, 7, 8]. In those theories, the (geometric) fitness [4, 5, 6] is generally maximised. This fitness measure equals the logarithm of the fold change in the number of organisms divided by the number of environmental periods. Organisms that produced most offspring have maximal geometric fitness. These theories evaluate the fitness contributions of phenotypic strategies, ranging from random, bet-hedging strategies to deterministic, responsive strategies. This work contributes to this growing body of theories.

Some phenotypic adaptation mechanisms are based on chance and are independent of environmental conditions. In such bet-hedging strategies [9], cells can switch randomly between alternative phenotypic states. This ensures the existence of subpopulations that are maladapted to the current condition, but are prepared for different future conditions. For example, slow-growing, stress-tolerant cells (so-called persister cells) and fast-growing, stress-sensitive cells have been discovered in microbial cell populations [10, 11, 12, 13]. One explanation is that the fast growing cells determine the current fitness, while the persister cells are 'insurance policies', guaranteeing survival when conditions suddenly become harsh and extinction-threatening. Together, they maximise the geometric fitness of the population in conditions that fluctuate between benign and existence-threatening conditions.

Close-to-deterministic adaptation strategies also exist, where cells perceive and transmit a signal to infer the state of the environment and adapt their phenotype accordingly. The omnipresence of two-component signalling circuits across microorganisms [14] indicates the importance of this mode of adaptation. However, inevitable molecular stochasticity in signalling circuits can cause random mismatches between the phenotype and the environmental

state, leading to fitness losses.

Since the entire range of phenotypic behaviours is found amongst microorganisms, and because they are great experimental systems for physiology and evolution, microbiology is the perfect playground for testing predictions and improving evolutionary theories. We therefore limit our theory to microbial phenotypic adaptation.

When bet-hedging adaptation strategies are more evolutionarily successful than sensing strategies has been a long standing question in evolutionary biology. Bet-hedging has been predicted to be evolutionarily advantageous under at least two conditions [9]: In slowly-changing, mild environments where sensing machinery would rarely provide an evolutionary benefit for long periods of time, and in environments that change quickly into extinction-threatening states where responsive adaptation would be too slow [1, 2, 3, 4, 6, 9]. In all other cases, responsive signalling strategies are favoured.

What remains poorly understood is what the optimal accuracy of signalling systems should be. Especially since experiments indicate that single cells have a remarkably low capacity for accurate tracking of environmental signals [15, 16, 17, 18, 19] below what is predicted to be possible from theory. Intuitively, one would expect that more accurate signalling improves evolutionary success as it reduces maladaptation. Accordingly, many theories are based on maximisation of the mutual information between signal and response [20, 21, 22]. However, in reality cells only appear to sense few external signals and in a noisy regime, leading to stochastic responses [23, 24, 25, 26]. Moreover, experiments report poor signalling capabilities of single cells; no more than 2 bits of mutual information between an external signal and an internal response have been found [16, 17, 18, 19]. Thus, either reliable signal transduction is not that important for evolutionary success, or it is very problematic for cells to achieve reliable signal transduction.

Reliable signal transduction might be difficult to achieve due to the inevitable randomness of molecular systems [27]. Increased expression of signalling proteins can make them more reliable [28], but requires a fitness-reducing diversion of resources from growth processes. Only if signalling activity leads to a net fitness increase will signalling systems evolve. Therefore, expression of idle signalling systems is generally fitness reducing in the absence of the signal and, moreover, genes encoding signalling machinery can then accumulate random mutations that reduce signalling performance. In addition, many signalling systems rely on signal receptors in the mem-

brane and their expression is at the expense of the nutrient importers, which reduces fitness too. These considerations may explain why cells express so few signalling systems; *E. coli* can grow, for instance, on hundreds of carbon sources, but it only has a handful of one and two-component sensing systems dedicated to nutrient sensing. However, these considerations do not explain why exploited signalling systems display such remarkably low mutual information values between their signal and their response. Answering this question is the goal of this paper.

In information-theoretic studies on cellular signalling, maximisation of mutual information between a signal and a response variable is considered beneficial for cells. We argue that this does not capture all situations. Imagine a flat relationship between the fitness value and the response variable versus one that is sharply peaked. In the former case, infinite mutual information between signal and response variable would not improve fitness, while it would in the latter case. Another limitation is that most theories consider only one signal and one response variable. It is likely that cells integrate different signals to induce a single response. Imagine a fitness landscape as a function of several response variables that contains ridges and peaks. In such a landscape, mutual information maximisation with respect to one signal and one response variable is insufficient. Therefore, phenotypic adaptation strategies should be evaluated on the basis of fitness maximisation, considering all signals. Importantly, also 'latent signals' – those that are not sensed by the cell, but that are fitness enhancing if they would be sensed – should be considered. Our theory incorporates this novel idea.

Our theory indicates that a signal should be transduced as accurately as possible only when it contains nearly all information about the optimal response. When information about the optimal response is partially contained in signals that are not sensed, it is evolutionarily optimal to perceive the sensed signals noisily. In this situation, the only way to increase fitness is to sense more signals. This suggests that cells should sense as many signals as possible, which might not be feasible as cells need to evolve separate signalling machinery for each signal. If those signals occur infrequently, and the signalling systems are generally idle, selection and drift will lead to their loss. Whether or not a cell evolves signalling circuits depends on the fitness gain that can be achieved by sensing this individual signal: when this fitness gain is higher than the fitness cost of having the sensing machinery, the machinery should evolve. This is why it can occur that it is optimal to sense individual signals noisily: the optimal accuracy of signal transduction decreases with

the cumulative latent information in all signals that did not reach the 'fitness threshold' for being sensed. When the ability to sense one signal is lost, the other sensed signals should not be perceived more, but less accurately. This is because, as the total available information about the optimal behaviour decreases, the cell should trust the available information less, and gamble more. This is the central idea of this paper, which we shall substantiate with theory.

The main insight from our work is that we can expect to find microorganisms in nature with poor signalling capacities because they live in an environment that they perceive only partially. They behave 'semi-blindly' not because they cannot do better, but because natural selection leads to maximisation of fitness by diverting biosynthetic resources to growth processes, at the expense of expressing signalling systems that either sense infrequent signals or those with moderate to low information about the optimal behaviour.

2. Results

2.1. The geometric fitness model

We aim to investigate fitness maximisation of an isogenic population in the presence of fluctuating signals. All single cells sense a signal s_s that induces the expression of a response protein with concentration p (figure 1A). In addition to the sensed signal, 'latent' signals, $s_{l1}, ..., s_{ln}$, exist that are not sensed by cells. They are the elements of the vector $\vec{s_l}$. All signals, both sensed and latent, determine the optimal protein concentration $p_o(s_s, \vec{s_l})$ that maximises fitness.

We use the following fitness function, which reaches its maximal value $f_o(s_s, \vec{s_l})$ at $p = p_o(s_s, \vec{s_l})$,

$$f(p; p_o(s_s, \vec{s_l}), f_o(s_s, \vec{s_l})) = f_o(s_s, \vec{s_l}) \exp\left\{-\frac{1}{2}\frac{(p - p_o(s_s, \vec{s_l}))^2}{\omega^2}\right\}. \quad (1)$$

The fitness reduction, due to a deviation of the protein concentration p from its optimal value p_o is determined by the width of the "fitness-landscape" ω (figures 1A and 1C). The interpretation of the fitness equation is explained in figure 1C.

To keep the model analytically tractable, we assume in addition that: (1) the optimum value p_o is a nonlinear function of s_s plus a weighted sum of all

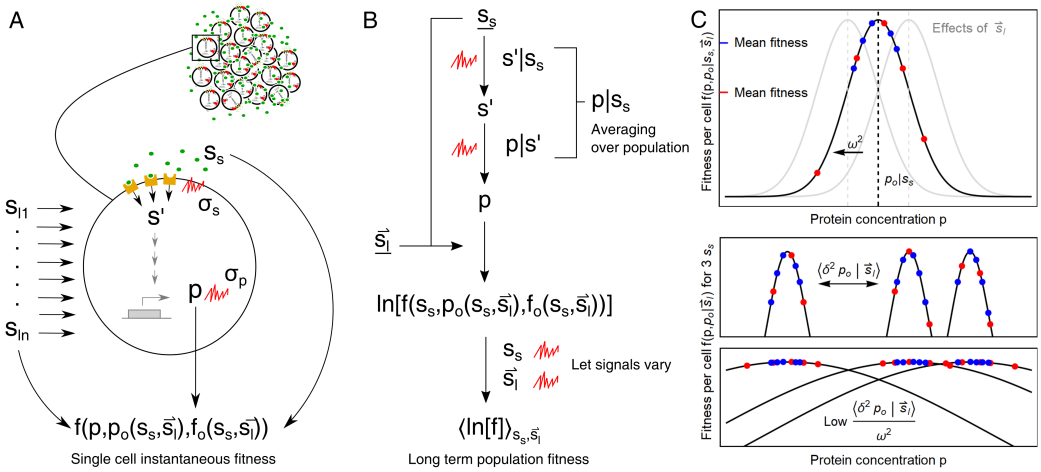

Figure 1: **Biological situation, methodological workflow, and intuition for the fitness function. A.** Evolution acts on an (isogenic) population. An individual cell senses an environmental signal s_s and estimates its concentration to be s'. In this process noise is added, this variance is σ_s^2 (to distinguish between two types of variances, variances denoted by σ^2 are intracellular variances (e.g. of a protein concentration) and variances denoted by $\langle \delta^2 x \rangle$ are extracellular variances (e.g. of an external signal concentration). The signal is transduced and eventually leads to protein concentration p, which contains noise on its own; with variance σ_p^2. The concentration p leads to a certain fitness of the cell depending on how far it is from the optimum p_o following the fitness function $f(p, p_o; f_o, \omega)$. The optimal protein concentration p_o is a function of the sensed signal s_s and all latent signals $s_{l1}, ..., s_{ln}$. Finally, f_o determines the fitness (growth rate) in the optimum. **B.** Under a fixed environment we use $s'|s_s \sim \mathcal{N}(s_s, \sigma_s^2)$ and $p|s' \sim \mathcal{N}(\alpha s', \sigma_p^2)$ to obtain $p|s_s \sim \mathcal{N}(\alpha s_s, \alpha^2\sigma_s^2 + \sigma_p^2)$, as shown in the appendix. The average fitness of the phenotypically heterogeneous population at one environment can now be determined and depends on s_s. Taking the logarithm and averaging over all signals (i.e. all environments) leads to the geometric mean of the fitness. **C.** The fitness of one cell is given by: $f(p, p_o(s_s, \vec{s_l}), f_o(s_s, \vec{s_l})) = f_o(s_s, \vec{s_l}) \exp\left\{-\frac{1}{2}\frac{(p-p_o(s_s,\vec{s_l}))^2}{\omega^2}\right\}$ where ω determines the width of the fitness function. The optimal protein concentration p_o is determined by the sensed signal s_s (black curve), but is also influenced by the latent signals $\vec{s_l}$ (gray curves). Considering two hypothetical populations of blue and red cells, where the blue cells are genetically superior in transducing signals, we see how the mean fitness of the population in one environment depends on the accuracy of estimating the optimum p_o. However, depending on the latent signals $\vec{s_l}$, the optimum and thus the mean fitness can still change. The middle plot shows how, under fixed $\vec{s_l}$, the variance in optima determines how far the peaks of the fitness curves lie apart. The bottom plot illustrates how inaccurate signal transduction will have fewer fitness consequences when the variance in optima is small relative to the width of the fitness curve.

latent signals $\vec{s}_l$ such that

$$p_o(s_s, \vec{s}_l) = p_o(s_s) + \sum_{i=1}^{n} \alpha_i s_{li}; \tag{2}$$

(2) all latent signals in $\vec{s}_l$ are independent, i.e. they do not covary, and are normally distributed; and (3) the means of $\vec{s}_l$ are set to 0.

Given these assumptions, following the methodology explained in figure 1B, we obtain for the geometric mean fitness of a population of isogenic cells (also see the appendix),

$$\begin{aligned} G = &\ln\left[f_o\right] - \frac{1}{2}\ln\left[\frac{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right] \\ &- \frac{1}{2}\frac{\int\left(\alpha s_s - p_o(s_s)\right)^2 h(s_s)ds_s + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}. \end{aligned} \tag{3}$$

The first term $\ln[f_o(s_s, \vec{s}_l)]$ equals the maximal geometric fitness. The last two terms are fitness costs. The first captures the cost of noisy signalling and quantifies the fitness reduction due to protein fluctuations that lead to an average deviation from the optimal concentration. The total variance of the protein fluctuations equals $\sigma_p^2 + \alpha^2\sigma_s^2$. The second cost term captures a distance from optimality due to two suboptimal (deterministic) effects: i. the relationship $p(s_s, , \vec{s}_l) = \alpha s_s$ might deviate from the optimal relation $p_o(s_s, \vec{s}_l)$ and ii. not sensing of latent signals that are informative about the optimal value of $p_o(s_s, \vec{s}_l)$, leads to a fitness loss too, in terms of the variances of these latent signals.

The last cost term in equation 3 contains $\int(\alpha s_s - p_o(s_s))^2 h(s_s)ds_s$, which is reminiscent of a mean squared error (MSE) of an estimator. Its occurrence suggests that evolution can be interpreted as minimising the distance of a cell's behaviour to its optimal behaviour. It is debatable whether this term can be made small by natural selection. For this to occur, the dependency $p_o(s_s)$ should not become a too complex function, as otherwise cells would not be able to approximate it by a steady-state input-output relation of a molecular circuit. Considering that we aim to investigate optimal sensing cells, we assume that $\alpha s_s \approx p_o(s_s)$ for any s_s. Therefore, the equation for the geometric mean fitness of sensing cells in the presence of multiple signals, both latent and observed, reduces to

$$G_s = \ln\left[f_o\right] - \frac{1}{2}\ln\left[\frac{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right] - \frac{1}{2}\frac{\sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}. \tag{4}$$

The final term in equation 4 is the fitness cost of varying latent signals. It quantifies how much fitness-influencing uncertainty remains about the environment, caused by the influence of all latent signals $\vec{s}_l$ on the optimum p_o. It is dependent on the width of the fitness function and the internal variance. This fitness cost can be reduced by increased protein fluctuations, i.e. of $\sigma_p^2 + \alpha^2\sigma_s^2$ by bet-hedging or 'noisy sensing' strategies, in agreement with previous results [1, 2, 3]. This can also be shown by differentiating equation 4 with respect to the normalised internal variance and solving for its optimum value, division by ω^2 then gives,

$$\left(\frac{\sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right)^{opt} = \frac{\sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle}{\omega^2} - 1; \tag{5}$$

indicating that the relative internal variance should only exceed zero when the relative uncertainty in the environment exceeds 1 (figure 2A) then noisy signalling pays off due uncertainty in the environment. The geometric mean fitness of a cell is more sensitive to the relative uncertainty about its environment than to its own relative internal variance (figure 2A).

Thus, an optimal sensing cell, which senses not all signals that are informative about the optimal behaviour, will still have to be noisy to overcome fitness variation due to the influences of all latent, not-sensed signals. Such a cell would therefore not follow a deterministic, pure sensing strategy, but allow for some bet-hedging behaviour; indicating that the mutual information between s_s and p should not always be maximised. This is an important insight from our theory that we will explore further.

2.2. Sensing vs non-sensing cells

The geometric mean fitness of non-sensing cells can be also derived from equation 3. First, σ_s^2 is set to 0; since in its absence, sensing cannot induce noise. Subsequently, we deduce that for an optimal, non-sensing cell $\langle p \rangle = \alpha s_s$ (see appendix). Finally, for a fair comparison to sensing cells, we have to consider the bias for non-sensing cells to be negligibly small such that $\langle p \rangle \approx \langle p_o(s_s) \rangle$. Integration of equation 3 over s_s under these conditions gives for geometric fitness the following relationship

$$G_{ns} = \ln\left[f_o\right] - \frac{1}{2}\ln\left[\frac{\omega^2 + \sigma_p^2}{\omega^2}\right] - \frac{1}{2}\frac{\langle \delta^2 p_o(s_s) \rangle + \sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle}{\omega^2 + \sigma_p^2}. \tag{6}$$

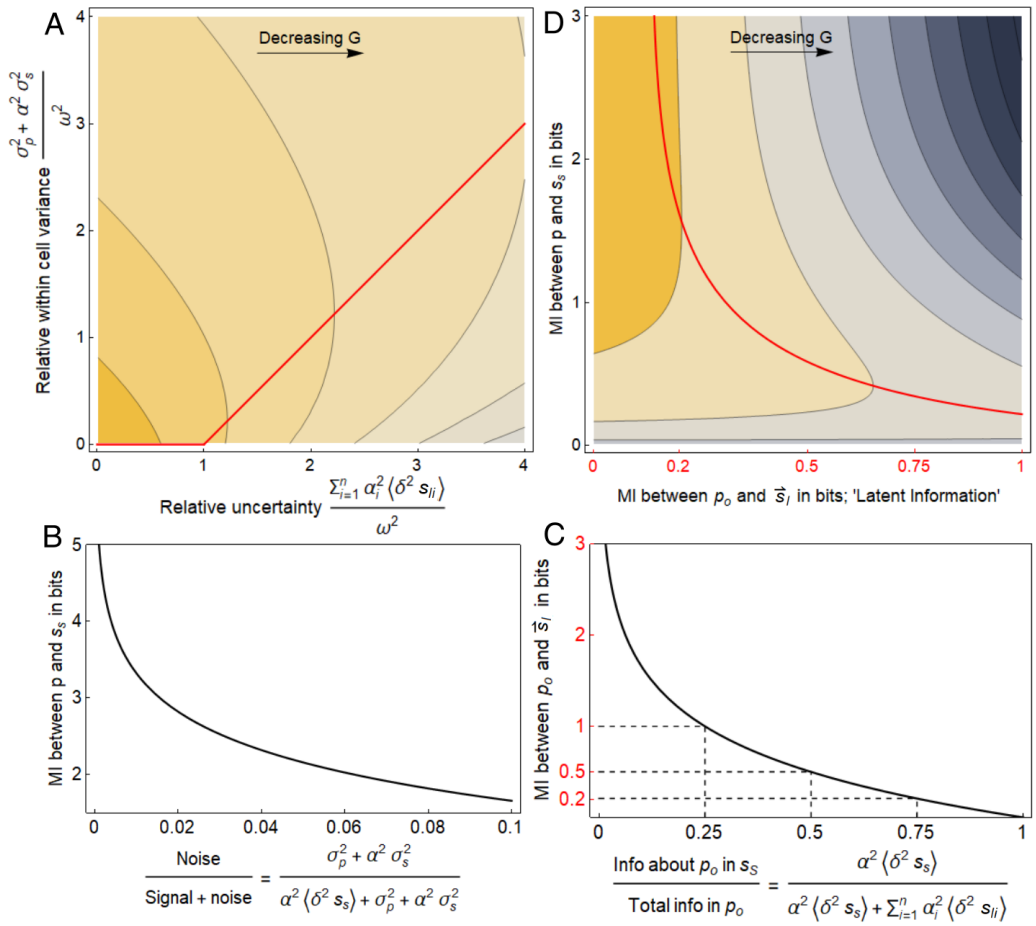

Figure 2: **Mutual information, latent information and fitness of sensing cells. A.** Contour plot of the geometric mean fitness (G_s) as a function of the relative within-cell variance and relative uncertainty in the environment. The colour indicates G_s as a percentage of the maximum (given by $\ln[f_o]$) here it ranges from 90% (bottom left) to 30% (bottom right). The red line shows the optimal relative internal variance for a given relative uncertainty. **B.** Mutual information (MI) between p and s_s as a function of the fraction of internal variance (i.e. noise) over all variance in p. This gives an indication of the range of the MI between p and s_s given a certain noise level. Note that the x-axis ranges from 0 to 0.1. **C.** MI between p_o and $\vec{s}_l$ as a function of the propagated variance in p_o from s_s over the total variance in p_o. Since we are considering normal distributions, variance is proportional to mutual information. **D.** Contour plot of G_s as a function of the MI between the sensed signal s_s and the cell's response p and the MI between the optimum p_o and all latent signals $\vec{s}_l$ (i.e. the latent information). The colouring indicates G_s as a percentage of the maximum; ranging from 75% (top left) to -150% (top right), the red line is the optimal MI between p and s_s, and $\rho = 5$. The x-axis covers a much smaller range than the y-axis, it corresponds to the red y-axis in C.

The optimal, normalised internal variance for non-sensing cells is given by,

$$\left(\frac{\sigma_p^2}{\omega^2}\right)^{opt} = \frac{\langle \delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li}\rangle}{\omega^2} - 1 \tag{7}$$

This result is similar to the results of Bull [1] and Haccou & Iwasa [2]. The total variance in the environment, defined in those papers as a single parameter, equals $\langle \delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li}\rangle$ in our extended formalism. Cells should introduce noise when the variance in the environment is exceeds the the width of the cellular fitness function.

We can now evaluate the fitness difference between sensing and non-sensing cells, i.e. $G_s^{opt} - G_{ns}^{opt}$ for three different scenarios; pure sensing vs pure non-sensing (low total variance in the optimum), pure sensing vs bet-hedging (most variance in the optimum is caused by s_s) and noisy sensing vs bet-hedging (high variance in optima caused by both s_s and $\vec{s}_l$). Here 'pure' indicates the strategy that aims to minimise internal variance. In each of these comparisons (see appendix), we find that $G_s^{opt} - G_{ns}^{opt}$ is minimally 0, which only occurs when $\langle \delta^2 p_o(s_s)\rangle = 0$ or $\omega^2 \to \infty$. These limits are in agreement with biological intuition, since the only reason not to sense a signal, when sensing does not come at any cost, is when it contains no information about the optimum or it has no fitness consequence.

2.3. Mutual and latent information

In order to interpret the geometric mean fitness of sensing cells from an information theoretic perspective we express it in terms of mutual information (MI). First we define the MI between the signal s_s that is being sensed and the response of the cell p. This way of using MI is very common in the evaluation of the accuracy of signal transduction. We find for the MI, in bits, between the sensed signal s_s and the internal response p (see appendix) the following relation

$$I(p; s_s) = -\frac{1}{2}\log_2\left[\frac{\sigma_p^2 + \alpha^2\sigma_s^2}{\alpha^2\langle \delta^2 s_s\rangle + \sigma_p^2 + \alpha^2\sigma_s^2}\right]. \tag{8}$$

The argument of the logarithm is the fraction of all variance in p that is caused by internal variance (i.e. noise). The MI depends on this fraction as shown in figure 2B, which indicates that when it decreases below 10%, the relevant range of MI is approximately 1-5 bits. The relevant range of MI as a function of the internal noise depends on both the type of signal transduction

and the choice of the input distribution. For instance, when two modes of a bimodal input distribution are perfectly separated and equally likely, the MI increases with 1 bit, effectively doubling the number of perceived states.

We will now introduce the in-our-eyes crucial information measure for the fitness of sensing cells. In contrast to what is commonly done, we define an MI term between the optimum and the latent signals. From the cell's perspective, this measure can be interpreted as 'latent information',

$$I(p_o; \vec{s}_l) = -\frac{1}{2}\log_2\left[\frac{\alpha^2\langle\delta^2 s_s\rangle}{\alpha^2\langle\delta^2 s_s\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}\right]. \tag{9}$$

The argument of the logarithm captures the fraction of information on the optimum that the cell retrieved (figure 2C). The relevant range of the latent information can be very low. For instance, when the sensed signal s_s contains 25% of the total information about the optimum, the latent information is 1 bit. When s_s contains 75% of all information on the optimum, the latent information is only 0.2 bit (figure 2C). An informative reference point to keep in mind is that at 0.5 bit latent information, s_s provides 50% of all information on the optimum p_o. Equation 9 indicates that the cell can reduce latent information by sensing more signals. By sensing an extra signal this signal's variance is added to the numerator in equation 9, decreasing the latent information. We note that latent information is not reduced by sensing an already sensed signal more accurately.

2.4. Maximal mutual information is not always optimal

We can write the geometric mean fitness of sensing cells (equation 4) in terms of the two mutual information measures. Before doing so, we define $\rho = \frac{\alpha^2\langle\delta^2 s_s\rangle}{\omega^2}$, the normalised variance in optima caused by the perceived signal s_s. The geometric mean fitness of a sensing population of isogenic cells equals

$$G_s = \ln[f_o] - \frac{1}{2}\ln\left[1 + \frac{\rho}{4^{I(p;s_s)} - 1}\right] - \frac{1}{2}\frac{\rho\,(4^{I(p_o;\vec{s}_l)} - 1)}{1 + \frac{\rho}{4^{I(p;s_s)}-1}}. \tag{10}$$

When we study G_s as a function of the mutual and latent information, given a certain value of ρ, we observe that the fitness decreases sharply in the direction of increasing latent information (figure 2D, note the low range on the x-axis).

Differentiating equation 10 with respect to the MI between p and s_s and setting it to 0 gives the optimal MI:

$$I(p;s_s)^{opt} = \frac{1}{2}\log_2\left[1+\frac{\rho}{\rho\left(4^{I(p_o;\vec{s}_l)}-1\right)-1}\right] \tag{11}$$

The optimal MI, $I(p;s_s)^{opt}$, given a certain amount of latent information, $I(p_o;\vec{s}_l)$, is plotted in red in figure 2D. The fact that there is an optimum for the MI means that cells should not maximise this term. This is exactly what a great part of previous research has focused on, as discussed in the introduction. The optimum is valid as long as $4^{I(p_o;\vec{s}_l)}-1>\frac{1}{\rho}$, which shows that as ρ increases there are lower values of $I(p_o;\vec{s}_l)$ for which an optimal number of bits MI between p and s_s exists. When this optimum does not exist, fitness is maximised as $I(p;s_s)$ approaches infinity, the fitness in this limit is given by:

$$\lim_{I(p;s_s)\to\infty} G_s = \ln\left[f_o\right] - \frac{1}{2}\rho\left(4^{I(p_o;\vec{s}_l)}-1\right) \tag{12}$$

This is the maximal fitness in the 'pure sensing' regime, where more MI always leads to higher fitness.

2.5. Distributed information leads to 'noisy sensing' of individual signals

An important conclusion can be drawn from figure 2D. It shows that for each given amount of latent information, there exists an optimal value of mutual information. This optimum converges to infinity as the latent information becomes very small. As discussed above, the only way for cells to reduce the latent information is to sense more signals. The red line in figure 2D shows that the mutual information should increase when the latent information decreases. This concerns not just the mutual information of p and s_s, but also of the newly sensed signal. This suggests that the total mutual information should increase as the cell reduces its latent information, but we do not know whether or not there is still an optimum for the mutual information per signal. Within our framework, the only reason to maximise the mutual information with one signal is that this signal contains almost all information about the optimal response of the cell.

2.6. High mutual information between one signal and a response seems rarely necessary

By distinguishing between the regimes where cells should maximise mutual information and where there exists an optimum, we can plot which

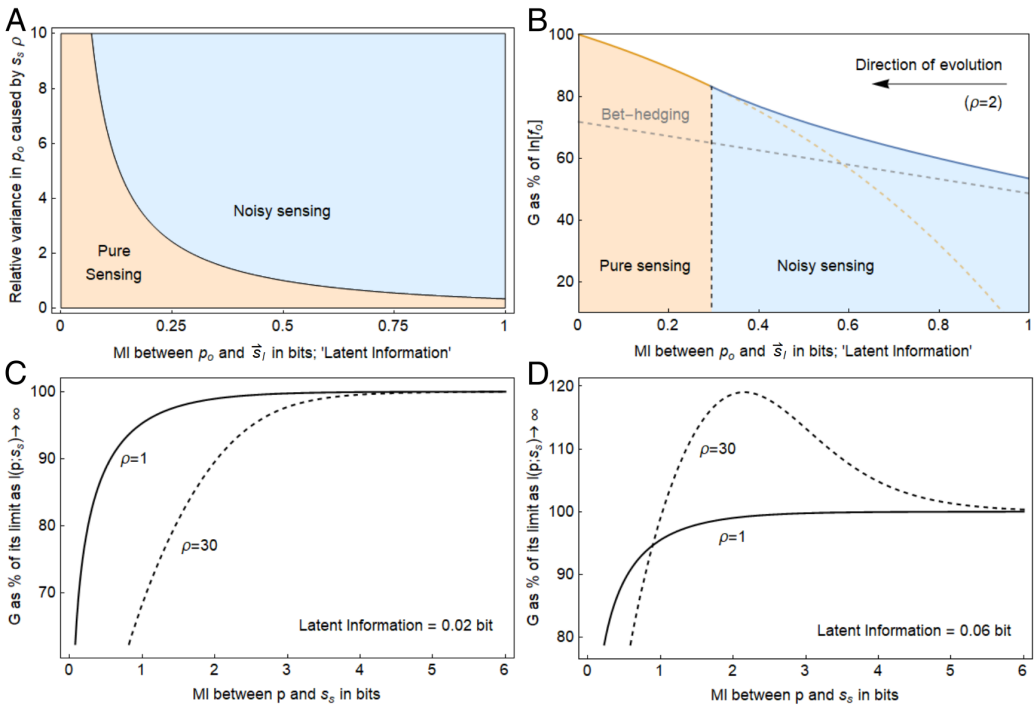

Figure 3: **Comparison of pure and noisy sensing strategies A.** Optimal strategies given a certain relative variance in the optimum p_o caused by s_s (ρ) and a certain number of bits of mutual information (MI) between p_o and all latent signals $\vec{s}_l$ (i.e. the latent information). In this plot the MI between p and s_s is optimal in each point, so ∞ in the pure sensing regime and $\frac{1}{2}\log_2\left[1+\frac{\rho}{\rho\,(4^{I(p_o;\vec{s}_l)}-1)-1}\right]$ in the noisy sensing regime. **B.** The geometric mean fitness G as a percentage of $\ln[f_o]$ (the maximal obtainable fitness) as a function of the MI between p_o and $\vec{s}_l$. Again the MI between p and s_s is optimal in each point. The dashed gray line shows the fitness of the optimal bet-hedging strategy in this situation. The direction of evolution is towards higher fitness, note that this can only be achieved by sensing more signals, as the total amount of information that is contained in one signal about the optimum is a given. **C.** G as a percentage of $\lim_{I(p;s_s)\to\infty} G$, as a function of the MI between p an s_s. The curve saturates quickly, even when ρ is increased drastically. The latent information (mutual information of p_o and $\vec{s}_l$) is 0.02 bit. **D.** G as a percentage of $\lim_{I(p;s_s)\to\infty} G$, as a function of the MI between p an s_s. As the latent information increases from 0.02 (in C) to 0.06 bit an optimum appears in the line where $\rho = 30$, meaning that in this situation cells should not maximise the MI between the sensed signal s_s and their response p.

strategy should be chosen for certain combinations of ρ and the latent information (figure 3A). Pure sensing is generally only preferred given low latent information. Another option seen in 3A, is when ρ becomes small; this is, however, an artificial situation as it means that the optimum would barely change. This solution exists because we do not consider costs to sensing. How low the latent information should become in order for pure sensing to become favourable depends on the value of ρ.

Evolution will drive the cell to reduced latent information values (figure 3B). However, cells can only achieve this by sensing more signals. In our model the cell would therefore start sensing all relevant signals, ending up in the pure sensing regime. In addition, as long as the variance in optima caused by the sensed signal ($\rho > 0$) remains, sensing is always preferred over bet-hedging. In reality however, cells will only sense signals that contain a sufficiently large and frequent enough fitness benefit, likely creating a lower bound on the latent information. As discussed, sensing machinery with low fitness benefit might be lost via selection and genetic drift; in figure 3B we see that this could happen at high latent information.

When one signal contains almost all information about the optimum, the mutual information should be maximised (figure 3C). We note, however, that even for high ρ the fitness curve saturates quickly, i.e. the amount of fitness gained per bit of mutual information quickly becomes very low. This is a consequence of two factors: (1) mutual information is an exponential measure, where the number of bits MI leads to 2^{MI} perceived states; (2) the usage of normal distributions leads to 4^{MI} instead of 2^{MI}, as can be seen in equation 10. In figure 3D we see how, when the latent information increases, the previously discussed optimum appears in the fitness curve. The exact latent information value when this occurs depends on the value of ρ.

3. Discussion

Maximisation of the mutual information between the input and the output of a signalling circuit has received a lot of theoretical attention. One of its advantages is that it can be applied without the use of a mechanistic, stochastic model of the signalling network. In one application, only three functions are needed: the input/output relation, the dependency of the output noise on the input, and the probability distributions of input values, i.e. of environmental states [20]. Maximisation of mutual information then allows for the prediction of one of those three functions from the two others.

Maximisation of mutual information is often rationalised by saying that it is a requirement for fitness maximisation. Our theory has shed doubt on this argument.

How precise the signalling machinery of a single microbe should track environmental signals depends on how important those signals are for its fitness. If latent signals exist, which would improve fitness if they would be sensed, then the accuracy with which sensed signals are transduced should be low. Thus, our theory predicts that the optimal mutual information between the sensed signal and the cell's response should generally not be maximal for fitness maximisation when latent signals occur.

It is likely that latent signals often occur. Firstly, because microorganisms display only a handful of signalling systems – in particular in the light of the huge number of nutrients that they can grow on. Secondly, expressing idle signalling systems reduces immediate fitness, because of biosynthetic resource consumption by non-growth promoting processes. Thirdly, idle, scanning signalling systems are evolutionary unstable; selection and drift would randomly mutate those unused systems. Thus, our theory sheds doubts on the relevance of maximisation of mutual information of signalling circuits. Natural selection for maximal geometric fitness in the presence of latent signals leads to optimal mutual information values that are not maximal.

Our approach does however have limitations, most of which are caused by our aim to create an analytically tractable model that has a general applicability. We have used non-truncated normal distributions, such that distributions of compound concentrations go to minus infinity. We assumed all variances to be independent of fluctuations in compound concentrations. Also, using normal distributions might by overly simplistic. Lastly, all assumptions of linearity might not always be in agreement with experimental data. However, by keeping our model analytically tractable, we obtained general qualitative insights into the fitness effects of phenotypic adaptation mechanisms from the consideration of a minimal model. When particular systems are of interest, models that are more mechanistic would provide additional, system-specific insight.

It has become clear in this work that the concept of mutual information should be used with care when it comes to quantifying cell performance. Firstly, biological performance should not be measured in terms of bits, but in terms of the fitness consequences of these bits. Secondly, it is difficult to distinguish characteristics of the input distribution from characteristics of the transduction system. This aspect of mutual information can lead to

a certain number of bits being perceived as low, but for the distribution of inputs that is being considered it might well be that it is close to all information that is contained in the input. For example, three bits mutual information does not sound as much, but when we consider figure 3C we see that it is actually relatively high for a normally distributed input. Thirdly, we show that the commonly used mutual information measure, the mutual information between a signal and a cell's response, is not the relevant measure when considering its fitness consequences. Together this does not mean that mutual information cannot be used in cell biology, it only means that one has to look at what is relevant for fitness. We have shown that in cellular adaptation there is a very relevant mutual information term, which is the mutual information between all latent signals and the optimum, the latent information. Thus, it is not about how well the cell can perceive one aspect of the environment, but about how the whole environment determines the optimal behaviour. Using this perspective, it became clear that only signals that contain nearly all information on the optimal behaviour should be sensed as accurately as possible.

Our model implies that cells should improve their fitness by sensing as many informative signals as are available. In reality, we see that cells only sense a few signals. Apart from the possibility of flux-based regulation (e.g. catabolite repression [29, 30]), this is likely caused by the fact that to obtain specific information on individual signals, a cell needs separate signalling machinery for each of these signals. Whether or not a cell evolves sensing machinery for a particular signal depends on the fitness gain that can be achieved by sensing this individual signal. When this fitness gain is higher than the fitness cost of having the sensing machinery, the machinery should evolve. Even when the fitness cost of having sensing machinery is very low, sensing machinery for signals with low or infrequent fitness gain will be lost due to genetic drift. So the signals that are being sensed by microbes must have exceeded the 'fitness threshold', having sensing machinery for these signals is evolutionarily beneficial. The accuracy with which these signals should be sensed depends on the cumulative information in all signals that did not reach the fitness threshold, and are thus not being sensed, i.e. the latent information. This leads to the somewhat counter-intuitive conclusion that when the ability to sense one signal is lost, the other sensed signals should be perceived less accurately – not more. As the total available information about the optimal behaviour decreases, the cell should trust the available information less, and gamble more. In conclusion: cells will not

evolve signalling systems for signals that provide a low or infrequent fitness gain, therefore they cannot precisely know the optimal behaviour. In this situation it is fitness reducing to sense other signals too accurately, so noisy signalling systems will evolve.

This work contributes to a better understanding of optimal phenotypic adaptation strategies and of the use of information theoretic concepts. By considering multiple signals in the light of their fitness consequence we were able to show that not the mutual information between one signal and one response is what is crucial to cells, but that the latent information is what ultimately determines evolutionary success.

4. Acknowledgements

We would like to thank dr. R. Planqué, Riccardo Muolo and Daan de Groot for their time and valuable comments during our weekly discussions, and Bram van de Putte, Tom Clement and Joanne Preuter for their support and insights over many cups of coffee.

Appendix A. Model derivations

For ease of comparison this appendix is structured in the same order as the results section.

Appendix A.1. The geometric fitness model

To determine the fitness of the population we start by considering the fitness of a single cell. We assume that a maximal fitness exists for each cell. In our case, the optimal behaviour, and deviation from this state, is determined by the protein concentration p. The sensing cell bases its concentration of p on an externally sensed signal with bulk concentration s_s. Since the cell infers s_s by sensing single molecules it will not know s_s exactly. It will make an estimate, s'. We assume that in a single cell $p|s'$ is normally distributed with a mean $\langle p|s'\rangle$ and a variance σ_s^2. We assume $\langle p|s'\rangle$ to be a linear function of s' such that $\langle p|s'\rangle = \alpha s'$. The variance σ_s^2 can be expressed in terms of signal concentration, diffusion constant, sample number and total time of sampling [31, 32]; we will not go into this further in this work and simply summarise these terms in the notation σ_s^2.

The optimal concentration of p is given by $p_o(s_s, \vec{s_l})$, which is a deterministic function of the sensed signal s_s and of all, so-called, 'latent signals'

whose values are the elements of the vector $\vec{s}_l$. A latent signal can be any measurable physical quantity, e.g. a nutrient concentration, in the cell's environment that the cell is not sensing, but that could be sensed by it, were the cell to change its genotype and evolve a sensing circuit. The influence of $\vec{s}_l$ on the optimal protein concentration is therefore defined as the concentration of p the cell should have had, when it would have had the required signalling circuits for estimation of the concentrations of these latent signals.

We define the fitness of one cell as follows,

$$f(p, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l)) = f_o(s_s, \vec{s}_l) \exp\left\{-\frac{1}{2}\frac{(p - p_o(s_s, \vec{s}_l))^2}{\omega^2}\right\}. \tag{A.1}$$

The function $f_o(s_s, \vec{s}_l)$ determines the maximal fitness when $p = p_o(s_s, \vec{s}_l)$. It is a deterministic function of s_s and $\vec{s}_l$. The parameter ω sets the width of the fitness function, the higher its value the more gentle the environment is (meaning optimality deviations have smaller fitness consequences), and vice versa. Equation A.1 is analogous to $\frac{N(p,t+\Delta t_i,s_s,\vec{s}_l)}{N(p,t,s_s,\vec{s}_l)} = e^{\mu_i(p,s_s,\vec{s}_l)\Delta t}$, it gives the fold change in number of organisms, N, at one value of p, in environment i, which lasts for Δt time, during which the organisms grow at specific growth rate $\mu_i(p, s_s, \vec{s}_l)$. Note we define a fixed time scale, Δt, after which we determine the fold change in the number of microorganisms; this time scale can be chosen small.

To determine the average growth of the population of a single genotype in a fixed environment we need to average over its phenotypes having different p values. We first determine the expected distribution of p over the whole population. We know $p|s' \sim \mathcal{N}\left(\alpha s', \sigma_p^2\right)$, and over the whole population $s'|s_s \sim \mathcal{N}\left(s_s, \sigma_s^2\right)$, corresponding to the probability density functions $g_{p|s'}(p|s')$ and $h_{s'|s_s}(s'|s_s)$ respectively. For the probability density function $g_{p|s_s}(p|s_s)$ we obtain that $p|s_s \sim \mathcal{N}\left(\alpha s_s, \sigma_p^2 + \alpha^2\sigma_s^2\right)$, i.e.

$$\begin{aligned} g_{p|s_s}(p|s_s) &= \int g_{p|s'}(p|s') h_{s'|s_s}(s'|s_s) ds' \\ &= \frac{1}{\sqrt{2\pi\left(\sigma_p^2 + \alpha^2\sigma_s^2\right)}} \exp\left\{-\frac{1}{2}\frac{(p - \alpha s_s)^2}{\sigma_p^2 + \alpha^2\sigma_s^2}\right\}. \end{aligned} \tag{A.2}$$

We can calculate the average fitness of the population over all concentrations p (all phenotypes) under fixed s_s and $\vec{s}_l$ by multiplying $f(p, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l))$

(eq. A.1) with $g(p|s_s)$ and integrating over p:

$$\begin{aligned}
&\langle f(p, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l))\,\rangle_p \\
&= \int f(p, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l)) g(p|s_s) dp \\
&= \frac{f_o(s_s, \vec{s}_l)\,\omega}{\sqrt{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}} \exp\left\{ -\frac{1}{2} \frac{(\alpha s_s - p_o(s_s, \vec{s}_l))^2}{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2} \right\} \\
&= f(s_s, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l))
\end{aligned} \tag{A.3}$$

This is the average fitness of the population at a fixed environment. What we have done so far is average the fitness in one environment over all phenotypes to obtain the fitness of the genotype in that environment i.

Next, we average the fitness over a sequence of environments. First we take the natural logarithm of equation A.3 and then calculate its expected value over s_s and $\vec{s}_l$. The equation we obtain gives the geometric mean of the population's fitness over all environments;

$$\begin{aligned}
G =& \langle \ln\left[f(s_s, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l))\right]\rangle_{s_s, \vec{s}_l} \\
=& \int\int \ln\left[f(s_s, p_o(s_s, \vec{s}_l), f_o(s_s, \vec{s}_l))\right] k_{s_s}(s_s) ds_s k_{\vec{s}_l}(\vec{s}_l) d\vec{s}_l \\
=& \langle \ln\left[f_o(s_s, \vec{s}_l)\right]\rangle_{s_s, \vec{s}_l} - \frac{1}{2} \ln\left[\frac{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right] \\
&- \frac{1}{2} \frac{\int\int (\alpha s_s - p_o(s_s, \vec{s}_l))^2 k_{s_s}(s_s) ds_s k_{\vec{s}_l}(\vec{s}_l) d\vec{s}_l}{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2}
\end{aligned} \tag{A.4}$$

What we have done now is equivalent to $\frac{1}{T}\sum_{i=1}^{N} \ln\left[e^{\mu_i, \Delta t_i}\right] = \langle \mu \rangle$, where the summation and division by $T = N\Delta t$ is used to average over all environments and $\frac{N_i \Delta t}{T} = p_i$ is the probability of an environment, with N_i as the number of times environment i occurred for duration Δt. We write the first term in equation A.4 as $\langle \ln\left[f_o\right]\rangle$, the mean maximal fitness over all environments. To simplify the third term of equation A.4 we make three assumptions: (1) the optimum is a function of s_s plus a weighted sum of all latent signals $\vec{s}_l$ such that;

$$p_o(s_s, \vec{s}_l) = p_o(s_s) + \sum_{i=1}^{n} \alpha_i s_{li} \tag{A.5}$$

(2) all latent signals in $\vec{s}_l$ are independent (they do not covary) and are normally distributed; (3) the means of each s_l are 0, thus the vector $\vec{s}_l$ contains the values of the latent signals in terms of a deviation from their means. Substituting $p_o(s_s, \vec{s}_l)$ gives for the numerator of the third term:

$$\int\int\left(\alpha s_s - p_o(s_s) - \sum_{i=1}^{n}\alpha_i s_{li}\right)^2 k_{s_s}(s_s)ds_s k_{\vec{s}_l}(\vec{s}_l)d\vec{s}_l \tag{A.6}$$

We assume that the cell's sensing machinery is optimised such that $\alpha s_s \approx p_o(s_s)$. The term is now independent of s_s, and considering all, n, latent signals are independent we can write them as separate integrals:

$$\int\cdots\int\left(\sum_{i=1}^{n}\alpha_i s_{li}\right)^2 \prod_{i=1}^{n} k_{s_{li}}(s_{li})ds_{li} \tag{A.7}$$

The square of the sum can be written as the sum of each term squared plus all their crossproducts:

$$\int\cdots\int\left(\sum_{i=1}^{n}\alpha_i^2 s_{li}^2 + \sum_{i=1}^{n}\sum_{j=1 j\neq i}^{n}\alpha_i\alpha_j s_{li}s_{hj}\right)\prod_{i=1}^{n} k_{s_{li}}(s_{li})ds_{li} \tag{A.8}$$

An integral of a sum can be written as the sum of integrals, α_i does not depend on s_{li} and can thus also be taken out of the integral, so we obtain:

$$\sum_{i=1}^{n}\alpha_i^2\langle s_{li}^2\rangle + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\alpha_i\alpha_j\langle s_{li}s_{hj}\rangle \tag{A.9}$$

The covariances and the means were assumed to be 0 so equation A.9 is equivalent to $\sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle$, inserting this result in equation A.4 and rearranging the second term, we obtain for the geometric mean fitness (i.e. the mean growth rate) of sensing cells:

$$G_s = \langle\ln[f_o]\rangle - \frac{1}{2}\ln\left[1 + \frac{\sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right] - \frac{1}{2}\frac{\sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}{\omega^2 + \sigma_p^2 + \alpha^2\sigma_s^2} \tag{A.10}$$

corresponding to equation 4 in the main text.

To obtain the optimal internal noise we differentiate equation A.10 with respect to the relative internal variance, $\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}$, and solve for the optimum:

$$\left(\frac{\sigma_p^2 + \alpha^2\sigma_s^2}{\omega^2}\right)^{opt} = \frac{\sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}{\omega^2} - 1 \tag{A.11}$$

This equation tells us internal variance is beneficial when $\frac{\sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}{\omega^2} > 1$; thus when the variance in the optimal protein level, due to not knowing the values of the latent signals, exceeds the width parameter of the fitness function it is optimal for cells to display noisy signalling. We expect this to be often the case, because: i. the sensing capacity of cells appears very low and ii. having sensing systems idle for long periods leads to a fitness costs that would eventually lead to the loss of those signalling systems.

Appendix A.2. Sensing vs non-sensing cells

To determine the geometric mean fitness of non-sensing cells we start with equation A.4. For non-sensing cells σ_s^2 equals 0 and $\alpha s_s = \langle p\rangle$; since there is no relationship between p and the signal, and we have already averaged over all values of p in equation A.3. We obtain for the fitness of a non-sensing cell

$$\begin{aligned} G_{ns} =& \langle \ln\left[f_o\right]\rangle - \frac{1}{2}\ln\left[1+\frac{\sigma_p^2}{\omega^2}\right] \\ &- \frac{1}{2}\frac{\int\int\left(\langle p\rangle - p_o(s_s,\vec{s}_l\,)\right)^2 k_{s_s}(s_s\,)ds_s k_{\vec{s}_l}(\vec{s}_l)d\vec{s}_l}{\omega^2+\sigma_p^2}. \end{aligned} \tag{A.12}$$

Again we insert the definition of $p_o(s_s,\vec{s}_l)$ from equation A.5 and inspect the numerator of the last term

$$\int\int\left(\langle p\rangle - p_o(s_s) - \sum_{i=1}^{n}\alpha_i s_{li}\right)^2 k_{s_s}(s_s\,)ds_s k_{\vec{s}_l}(\vec{s}_l)d\vec{s}_l \tag{A.13}$$

For now we will just consider the integration over s_s, expanding the squared term we obtain:

$$\begin{aligned} \int\Bigg(&\langle p\rangle^2 + p_o(s_s)^2 + \left(\sum_{i=1}^{n}\alpha_i s_{li}\right)^2 - 2\langle p\rangle p_o(s_s) \\ &- 2\langle p\rangle\sum_{i=1}^{n}\alpha_i s_{li} + 2p_o(s_s)\sum_{i=1}^{n}\alpha_i s_{li}\Bigg)k_{s_s}(s_s\,)ds_s \end{aligned}, \tag{A.14}$$

which equals

$$\begin{aligned} &\langle p\rangle^2 + \langle p_o(s_s)^2\rangle + \left(\sum_{i=1}^{n}\alpha_i s_{li}\right)^2 - 2\langle p\rangle\langle p_o(s_s)\rangle \\ &- 2\langle p\rangle\sum_{i=1}^{n}\alpha_i s_{li} + 2\langle p_o(s_s)\rangle\sum_{i=1}^{n}\alpha_i s_{li}. \end{aligned} \tag{A.15}$$

We assume, similar to the optimisation assumption for sensing cells, that the mean protein concentration is optimised such that $\langle p \rangle \approx \langle p_o(s_s) \rangle$. We obtain:

$$\begin{aligned} &\langle p_o(s_s)^2 \rangle - \langle p_o(s_s) \rangle^2 + \left(\sum_{i=1}^{n} \alpha_i s_{li} \right)^2 \\ &= \langle \delta^2 p_o(s_s) \rangle + \left(\sum_{i=1}^{n} \alpha_i s_{li} \right)^2 \end{aligned} \tag{A.16}$$

The second term in this equation still needs to be integrated over all latent signals $\vec{s}_l$. This is equivalent to the derivation from equation A.7 to A.9, so the result will be $\langle \delta^2 p_o(s_s) \rangle + \alpha^2 \langle \delta^2 s_{li} \rangle$. Inserting this result in equation A.12 we obtain:

$$G_{ns} = \langle \ln [f_o] \rangle - \frac{1}{2} \ln \left[1 + \frac{\sigma_p^2}{\omega^2} \right] - \frac{1}{2} \frac{\langle \delta^2 p_o(s_s) \rangle + \alpha^2 \langle \delta^2 s_{li} \rangle}{\omega^2 + \sigma_p^2}, \tag{A.17}$$

which corresponds to equation 6 in the main text. Analogous to equation A.11, the optimal, normalized internal variance for non-sensing cells is given by:

$$\left(\frac{\sigma_p^2}{\omega^2} \right)^{opt} = \frac{\langle \delta^2 p_o(s_s) \rangle + \sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle}{\omega^2} - 1 \tag{A.18}$$

This equation indicates that having internal noise is beneficial if $\frac{\langle \delta^2 p_o(s_s) \rangle + \sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle}{\omega^2} > 1$ which is more stringent than the previous condition because of added term $\langle \delta^2 p_o(s_s) \rangle$, which captures the uncertainty in the optimal value of p due to fluctuations in s_s.

As has been stated in the main text, there exists relevant comparisons between the sensing and non-sensing strategies; pure sensing vs pure non-sensing, pure sensing vs bet-hedging and noisy sensing vs bet-hedging. There exists no regime where noisy sensing and pure non-sensing would be preferred at the same time. Here we will subtract the optimal non-sensing strategy from the optimal sensing strategy for each of the three possible situations. First we will derive the four fitness formulas where the internal variance is optimal. Let us define $u = \frac{\sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle}{\omega^2}$ and $\rho = \frac{\langle \delta^2 p_o(s_s) \rangle}{\omega^2}$.

When $u \leq 1$ pure sensing is optimal so we have $\left(\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}\right)^{opt} = 0$, hence we obtain for the geometric mean fitness (starting from equation A.10):

$$\begin{aligned} G_s^{opt-pure} &= \ln[f_o] - \frac{1}{2}\ln\left[1+\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}\right] - \frac{1}{2}\frac{u}{1+\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}} \\ &= \ln[f_o] - \frac{1}{2}u \end{aligned} \tag{A.19}$$

For $u > 1$ we have noisy sensing with in the optimum $\left(\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}\right)^{opt} = u-1$ giving:

$$\begin{aligned} G_s^{opt-noisy} &= \ln[f_o] - \frac{1}{2}\ln\left[1+\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}\right] - \frac{1}{2}\frac{u}{1+\frac{\sigma_p^2+\alpha^2\sigma_s^2}{\omega^2}} \\ &= \ln[f_o] - \frac{1}{2}\ln[u] - \frac{1}{2} \end{aligned} \tag{A.20}$$

Now for the non-sensing strategies; when $u+\rho \leq 1$ we are in the pure regime where $\left(\frac{\sigma_p^2}{\omega^2}\right)^{opt} = 0$ leading to the optimal G:

$$\begin{aligned} G_{ns}^{opt-pure} &= \ln[f_o] - \frac{1}{2}\ln\left[1+\frac{\sigma_p^2}{\omega^2}\right] - \frac{1}{2}\frac{u+\rho}{1+\frac{\sigma_p^2}{\omega^2}} \\ &= \ln[f_o] - \frac{1}{2}u - \frac{1}{2}\rho \end{aligned} \tag{A.21}$$

For $u+\rho > 1$ the optimal non-sensing strategy is bet-hedging with $\left(\frac{\sigma_p^2}{\omega^2}\right)^{opt} = u+\rho-1$ which leads to:

$$\begin{aligned} G_{ns}^{opt-bh} &= \ln[f_o] - \frac{1}{2}\ln\left[1+\frac{\sigma_p^2}{\omega^2}\right] - \frac{1}{2}\frac{u+\rho}{1+\frac{\sigma_p^2}{\omega^2}} \\ &= \ln[f_o] - \frac{1}{2}\ln[u+\rho] - \frac{1}{2} \end{aligned} \tag{A.22}$$

Now we can calculate the fitness difference between the strategies in each possible regime, starting with $u+\rho \leq 1$ (and so $u \leq 1$ as well):

$$G_s^{opt-pure} - G_{ns}^{opt-pure} = \frac{1}{2}\rho \tag{A.23}$$

For $u + \rho > 1$ and $u \leq 1$ we get:

$$G_s^{opt-pure} - G_{ns}^{opt-bh} = \frac{1}{2}\ln[u+\rho] + \frac{1}{2} - \frac{1}{2}u \tag{A.24}$$

And finally, for $u > 1$ (so $u + \rho > 1$ as well):

$$G_s^{opt-noisy} - G_{ns}^{opt-bh} = \frac{1}{2}\ln[u+\rho] - \frac{1}{2}\ln[u] \tag{A.25}$$

It is clear that the minimum of equations A.23 to A.25 is 0 (considering all parameters have to be positive) and that this minimum is only reached when $\rho = 0$. In the case of equation A.24 this can only happen when $u = 1$ (and even then $u + \rho > 1$ strictly speaking does not hold when $\rho = 0$). Let us remember that, $\rho = \frac{\langle \delta^2 p_o(s_s)\rangle}{\omega^2}$ and $u = \frac{\sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li}\rangle}{\omega^2}$. From this we can conclude that, in our model, the only reasons not to sense are that the signal of interest s_s has no influence on the optimal response, or the width of the fitness function ω^2 is infinitely large (leading to $u = 0$ as well as $\rho = 0$). This is to be expected when sensing does not come at any cost.

Appendix A.3. Expressing noise in terms of mutual information

We first define the MI between the signal that is being sensed s_s and the response of the cell p. Following the definition of MI [33, 34] we obtain;

$$I(p; s_s) = \mathcal{H}(p) - \langle \mathcal{H}(p\,|s_s)\rangle_{s_s} \tag{A.26}$$

where $\mathcal{H}$ denotes the entropy. We know from equation A.2 that $p|s_s \sim \mathcal{N}\left(\alpha s_s,\, \sigma_p^2 + \alpha^2\sigma_s^2\right)$ and given $s_s \sim \mathcal{N}\left(\langle s_s\rangle,\, \langle \delta^2 s_s\rangle\right)$ we can obtain the probability density function of p:

$$\begin{aligned} h(p) &= \int h(p|s_s)h(s_s)ds_s \\ &= \frac{\exp\left\{-\frac{1}{2}\frac{(p-\alpha\langle s_s\rangle)^2}{\alpha^2\langle\delta^2 s_s\rangle+\sigma_p^2+\alpha^2\sigma_s^2}\right\}}{\sqrt{2\pi\left(\alpha^2\langle\delta^2 s_s\rangle + \sigma_p^2 + \alpha^2\sigma_s^2\right)}} \end{aligned} \tag{A.27}$$

Using the definition of the entropy of a normal distribution we insert the probability density functions of p and $p|s_s$ in equation A.26:

$$\begin{aligned} I(p;s_s) =& \frac{1}{2}\log_2\left[2\pi e\left(\alpha^2\langle\delta^2 s_s\rangle + \sigma_p^2 + \alpha^2\sigma_s^2\right)\right] - \frac{1}{2}\log_2\left[2\pi e\left(\sigma_p^2 + \alpha^2\sigma_s^2\right)\right] \\ =& -\frac{1}{2}\log_2\left[\frac{\sigma_p^2 + \alpha^2\sigma_s^2}{\alpha^2\langle\delta^2 s_s\rangle + \sigma_p^2 + \alpha^2\sigma_s^2}\right] \end{aligned} \tag{A.28}$$

The argument of the logarithm in equation A.28 is the fraction of all variance in p that is caused by internal variance (i.e. noise).

Appendix A.4. Defining latent information: mutual information between the optimum and all latent signals

We can derive the MI between the optimum p_o and all latent signals $\vec{s}_l$ in a similar manner as we have derived the MI between p and s_s above. This measure is what we call the 'latent information'. As p_o is the sum of independent normally distributed random variables, as shown in equation A.5, and having assumed the means of $\vec{s}_l$ to be 0, we obtain a normally distributed p_o with the following mean and variance respectively:

$$\langle p_o\rangle = \langle p_o(s_s)\rangle \tag{A.29}$$

$$\langle\delta^2 p_o\rangle = \langle\delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle \tag{A.30}$$

So its probability density function is given by:

$$h(p_o) = \frac{\exp\left\{-\frac{1}{2}\frac{(p-\langle p_o(s_s)\rangle)^2}{\langle\delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}\right\}}{\sqrt{2\pi\left(\langle\delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle\right)}} \tag{A.31}$$

It is clear that when $\vec{s}_l$ is fixed, containing all latent signals but not the sensed signal, the variances of the latent signals will be 0:

$$h(p_o|\vec{s}_l) = \frac{1}{\sqrt{2\pi\langle\delta^2 p_o(s_s)\rangle}}\exp\left\{-\frac{1}{2}\frac{\left(p-\langle p_o(s_s)\rangle\right)^2}{\langle\delta^2 p_o(s_s)\rangle}\right\} \tag{A.32}$$

Here the mean of the function is not conditional on $\vec{s}_l$ since all signals were assumed to be independent. Now we can make use of the probability density functions defined in equations A.31 and A.32 and the definition of the entropy of a normal distribution, to calculate the MI between p_o and $\vec{s}_l$:

$$
\begin{aligned}
I(p_o;\vec{s}_l) &= \mathcal{H}\left(p_o\right) - \langle \mathcal{H}\left(p_o\,|\vec{s}_l\,\right)\rangle_{\vec{s}_l} \\
&= \frac{1}{2}\log_2\left[2\pi e\left(\langle\delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle\right)\right] - \frac{1}{2}\log_2\left[2\pi e\left(\langle\delta^2 p_o(s_s)\rangle\right)\right] \\
&= -\frac{1}{2}\log_2\left[\frac{\langle\delta^2 p_o(s_s)\rangle}{\langle\delta^2 p_o(s_s)\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}\right]
\end{aligned}
\tag{A.33}
$$

We have assumed above that $\alpha s_s \approx p_o(s_s)$. This also means that $\langle\delta^2 p_o(s_s)\rangle \approx \alpha^2\langle\delta^2 s_s\rangle$. We use the latter notation for ease of comparison between the two MI terms. So we obtain for the MI between p_o and $\vec{s}_l$, i.e. the latent information:

$$
I(p_o;\vec{s}_l) = -\frac{1}{2}\log_2\left[\frac{\alpha^2\langle\delta^2 s_s\rangle}{\alpha^2\langle\delta^2 s_s\rangle + \sum_{i=1}^{n}\alpha_i^2\langle\delta^2 s_{li}\rangle}\right] \tag{A.34}
$$

Here the fraction inside the logarithm shows how much information the cell has on the optimum over all information that determines the optimum itself. By sensing an extra signal this signal's variance is added in the numerator in equation A.34, decreasing the latent information.

Appendix A.5. Writing the fitness of sensing cells in terms of mutual and latent information

Using the MI between p and s_s (equation A.28) the MI between p_o and $\vec{s}_l$ (equation A.34, the latent information), we can write the geometric mean fitness of sensing cells (equation A.10) in terms of these measures. We first rewrite equation A.28 to derive:

$$
\sigma_p^2 + \alpha^2\sigma_s^2 = \frac{\alpha^2\langle\delta^2 s_s\rangle}{4^{I(p;s_s)} - 1} \tag{A.35}
$$

And similarly from equation 9:

$$\sum_{i=1}^{n} \alpha_i^2 \langle \delta^2 s_{li} \rangle = \alpha^2 \langle \delta^2 s_s \rangle (4^{I(p_o;\vec{s}_l)} - 1) \tag{A.36}$$

Now we define $\rho = \frac{\alpha^2 \langle \delta^2 s_s \rangle}{\omega^2}$, which is a measure for the variance in optima caused by the perceived signal s_s, relative to the fitness width. Inserting our results from equations A.35 and A.36 in equation A.10 gives:

$$\begin{aligned} G_s =& \ln [f_o] - \frac{1}{2} \ln \left[1 + \frac{\alpha^2 \langle \delta^2 s_s \rangle}{\omega^2 (4^{I(p;s_s)} - 1)} \right] - \frac{1}{2} \frac{\alpha^2 \langle \delta^2 s_s \rangle (4^{I(p_o;\vec{s}_l)} - 1)}{\omega^2 + \frac{\alpha^2 \langle \delta^2 s_s \rangle}{4^{I(p;s_s)} - 1}} \\ =& \ln [f_o] - \frac{1}{2} \ln \left[1 + \frac{\rho}{4^{I(p;s_s)} - 1} \right] - \frac{1}{2} \frac{\rho \left(4^{I(p_o;\vec{s}_l)} - 1\right)}{1 + \frac{\rho}{4^{I(p;s_s)} - 1}} \end{aligned} \tag{A.37}$$

Differentiating equation A.37 with respect to the MI between p and s_s and setting it to 0 gives the optimal MI:

$$I(p; s_s)^{opt} = \frac{1}{2} \log_2 \left[1 + \frac{\rho}{\rho \left(4^{I(p_o;\vec{s}_l)} - 1\right) - 1} \right] \tag{A.38}$$

This optimum is valid as long as $4^{I(p_o;\vec{s}_l)} - 1 > \frac{1}{\rho}$, which shows that as ρ increases there are lower values of $I(p_o; \vec{s}_l)$ for which an optimal number of bits MI between p and s_s exists. When this optimum does not exist fitness is maximised as $I(p; s_s)$ approaches infinity, the fitness in this limit is given by:

$$\lim_{I(p;s_s) \to \infty} G_s = \ln [f_o] - \frac{1}{2} \rho \left(4^{I(p_o;\vec{s}_l)} - 1\right) \tag{A.39}$$

This is the maximal fitness in the 'pure sensing' regime, where more MI always leads to higher fitness.

References

[1] J. Bull, Evolution of phenoptypic variance, Evolution 41 (1987) 303–315.

[2] P. Haccou, Y. Iwasa, Optimal mixed strategies in stochastic environments, Theoretical Population Biology 47 (1995) 212–243.

[3] L. Wolf, O. K. Silander, E. van Nimwegen, Expression noise facilitates the evolution of gene regulation, eLife 4 (2015) 1–48.

[4] E. Kussell, S. Leibler, Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments, Science 309 (2005) 2075–2078.

[5] B. E. Sæther, S. Engen, The concept of fitness in fluctuating environments, Trends in Ecology and Evolution 30 (2015) 273–281.

[6] O. Rivoire, S. Leibler, The value of information for populations in varying environments, Journal of Statistical Physics 142 (2011) 1124–1166.

[7] M. Acar, J. T. Mettetal, A. van Oudenaarden, Stochastic switching as a survival strategy in fluctuating environments, Nature genetics 40 (2008).

[8] J. M. Pedraza, D. A. Garcia, M. F. Pérez-Ortiz, Noise, Information and Fitness in Changing Environments, Frontiers in Physics 6 (2018) 1–12.

[9] T. M. Norman, N. D. Lord, J. Paulsson, R. Losick, Stochastic Switching of Cell Fate in Microbes, Annual Review of Microbiology 69 (2015) 381–403.

[10] J. Bigger, Treatment of staphylococcal infections with penicillin by intermittent sterilisation, The Lancet 244 (1944) 497–500.

[11] K. Lewis, Persister cells, dormancy and infectious disease, Nature Reviews Microbiology 5 (2007) 48.

[12] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, Bacterial persistence as a phenotypic switch, Science 305 (2004) 1622–1625.

[13] A. Harms, E. Maisonneuve, K. Gerdes, Mechanisms of bacterial persistence during stress and antibiotic exposure, Science 354 (2016).

[14] A. M. Stock, V. L. Robinson, P. N. Goudreau, Two-component signal transduction, Annual review of biochemistry 69 (2000) 183–215.

[15] C. G. Bowsher, P. S. Swain, Environmental sensing, information transfer, and cellular decision-making, Current Opinion in Biotechnology 28 (2014) 149–155.

[16] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, Science (2011) 1204553.

[17] M. Coppey, A. N. Boettiger, A. M. Berezhkovskii, S. Y. Shvartsman, Nuclear trapping shapes the terminal gradient in the drosophila embryo, Current Biology 18 (2008) 915–919.

[18] S. Uda, T. H. Saito, T. Kudo, T. Kokaji, T. Tsuchiya, H. Kubota, Y. Komori, Y.-i. Ozaki, S. Kuroda, Robustness and compensation of information transmission of signaling pathways, Science 341 (2013) 558–561.

[19] M. Voliotis, R. M. Perrett, C. McWilliams, C. A. McArdle, C. G. Bowsher, Information transfer by leaky, heterogeneous, protein kinase signaling systems, Proceedings of the National Academy of Sciences 111 (2014) E326–E333.

[20] W. Bialek, Biophysics: searching for principles, Princeton University Press, Woodstock, Oxfordshire, 2012.

[21] L. A. Rhee Alex, Cheong Raymond, The application of information theory to biochemical signaling systems, Physical Biology 9 (2013).

[22] A. Levchenko, I. Nemenman, Cellular noise and information transmission, Current Opinion in Biotechnology 28 (2014) 156–164.

[23] T. J. Perkins, P. S. Swain, Strategies for cellular decision-making, Molecular Systems Biology 5 (2009).

[24] O. Feinerman, J. Veiga, J. R. Dorfman, R. N. Germain, G. Altan-Bonnet, Variability and robustness in t cell activation from regulated heterogeneity in protein levels, Science 321 (2008) 1081–1084.

[25] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, et al., Oscillations and variability in the p53 system, Molecular systems biology 2 (2006).

[26] A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov, U. Alon, Variability and memory of protein levels in human cells, Nature 444 (2006) 643.

[27] A. Eldar, M. B. Elowitz, Functional roles for noise in genetic circuits, Nature 467 (2010) 167.

[28] J. Paulsson, Summing up the noise in gene networks, Nature 427 (2004) 415.

[29] B. Görke, J. Stülke, Carbon catabolite repression in bacteria: many ways to make the most out of nutrients, Nature Reviews Microbiology 6 (2008) 613.

[30] V. Chubukov, L. Gerosa, K. Kochanowski, U. Sauer, Coordination of microbial metabolism, Nature Reviews Microbiology 12 (2014) 327–340.

[31] H. Berg, E. Purcell, Physics of chemoreception, Biophysical Journal 20 (1977) 193 – 219.

[32] C. C. Govern, P. R. ten Wolde, Fundamental limits on sensing chemical concentrations with linear biochemical networks, Physical review letters 109 (2012) 218103.

[33] C. E. Shannon, A mathematical theory of communication, Bell system technical journal 27 (1948) 379–423.

[34] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, 2012.

[35] K. Kaizu, W. De Ronde, J. Paijmans, K. Takahashi, F. Tostevin, P. R. Ten Wolde, The berg-purcell limit revisited, Biophysical journal 106 (2014) 976–985.

[36] A. Schwabe, T. R. Maarleveld, F. J. Bruggeman, Exploration of the spontaneous fluctuating activity of single enzyme molecules Anne, FEBS Letters (2013) 2744–2752.

[37] M. Kimura, The neutral theory of molecular evolution, Cambridge University Press, 1983.

[38] M. C. Whitlock, R. Buerger, Fixation of New Mutations in Small Populations, Evolutionary Conservation Biology (2004) 155–170.

[39] M. Lynch, B. Walsh, The origins of genome architecture, volume 98, Sinauer Associates Sunderland (MA), 2007.

[40] A. Becskei, B. B. Kaufmann, A. van Oudenaarden, Contributions of low molecule number and chromosomal positioning to stochastic gene expression, Nature genetics 37 (2005) 937.

[41] T. Stoeger, N. Battich, L. Pelkmans, Passive Noise Filtering by Cellular Compartmentalization, Cell 164 (2016) 1151–1161.

[42] C. S. Maxwell, P. M. Magwene, When sensing is gambling: An experimental system reveals how plasticity can generate tunable bet-hedging strategies, Evolution 71 (2017) 859–871.

[43] E. Libby, T. J. Perkins, P. S. Swain, Noisy information processing through transcriptional regulation., Pnas 104 (2007) 7151–6.

[44] R. Milo, R. Phillips, Cell biology by the numbers, Garland Science, New York, 2016.

[45] H. A. Orr, Fitness and its role in evolutionary genetics, Nature Reviews Genetics 10 (2010) 531–539.