Long title: The positive role of noise for information acquisition in biological signaling pathways

Short title: Noise can increase information acquisition in signaling pathways

Eugenio Azpeitia^{1,2,3}, Andreas Wagner^{1,2,4}

1. Department of Evolutionary Biology and Environmental Studies, University of Zürich, Zürich, Switzerland.

2. Swiss Institute of Bioinformatics, Lausanne, Switzerland.

3. Centro de Ciencias Matemáticas, UNAM, Morelia, México.

4. The Santa Fe Institute, Santa Fe, NM, USA.

Abstract

All living systems acquire information about their environment. At the cellular level, they do so through signaling pathways, which rely on interactions between molecules that detect and transmit the presence of an extracellular cue or signal to the cell's interior. Such interactions are inherently stochastic and thus noisy. In classical information theory, a noisy communication channel degrades the amount of transmissible information relative to a noise-free channel. For this reason, one would expect that the kinetic parameters that determine a pathway's operation minimize noise. We show that this is not the case under a wide range of biologically sensible parameter values. Specifically, we perform computational simulations of simple signaling systems, which show that a noisy molecular interaction dynamics is a necessary condition for information acquisition. Moreover, we show that optimal information acquisition, where a system reacts most sensitively to changes in the environment, can be obtained close to the maximal attainable level of noise in the

system. Our work highlights the positive role that noise can have in biological information processing.

Author summary

The acquisition of information is fundamental for living systems, because the decisions they take based on such information directly affect survival and reproduction. The molecular mechanisms used by cells to acquire information are signaling pathways. The molecular interactions of signaling pathways, such as the binding of a signal to a receptor, are by nature noisy. This is important, because noise disrupts information. Hence, to maximize the acquisition of information, signaling pathways should minimize the noise of their molecular interactions. Here we show that a noisy dynamic of the molecular interactions can improve the acquisition of information, and that the maximal capacity to acquire information can be obtained with a close-to-maximal level of noise in a signaling pathway. Thus, contrary to expectations, noise can improve the acquisition of information in living systems.

Introduction

Information about the environment is fundamental when living organisms make decisions that affect their survival and reproduction (1). For example, microbes adjust their growth in response to nutrient concentrations, animals flee in response to predators, and plants synthesize defense chemicals in response to herbivores.

At the cellular level, signaling pathways are the main molecular mechanism by which organisms acquire information. They typically detect the presence of a molecular signal or cue (2) about the environment through the binding of this molecule to a receptor.

Once the signal has been detected, a chain of intermediary events transmits this information to the cell's interior, where it ultimately regulates gene expression. Signaling pathways vary widely, including in their number of molecular interactions, signal and receptor affinities, the presence of feedback and feed-forward interactions, and the number of regulated genes (3,4). However, they all share some elementary processes, such as the reversible binding of molecules, which is necessary to detect a signal by a receptor, transmit its presence via effector molecules, for example through allosteric control of these molecules, and regulate gene expression through the binding of transcription factors to DNA.

Noise is present at all spatial and temporal scales of biological organization, from population dynamics to molecular interactions, including signaling pathways (5,6). The sources of noise in signaling pathways include random fluctuations in the concentration, movement, activity, and interactions of molecules (7–12). Classical information theory predicts that noise degrades the capability of a communication channel to transmit information. For example, in simple systems such as a binary or a symmetric information transmission channel, the maximum capacity of the channel to transmit information can only be realized in the absence of noise (13). In signaling pathways, noise transforms a stimulus, such as the concentration of a nutrient, into a distribution of outputs or responses. Any overlap between the response distributions produced by two different stimuli, such as two different signal concentrations, creates uncertainty about which stimuli produced which output (Fig 1a; 14). For this reason, noise also decreases the information acquisition capacity of signaling pathways.

Fig 1. Relationship of acquired information with both noise and output range. Example of two hypothetical stimuli that produce different but overlapping response distributions (green and blue distributions). The amount of information acquired at different levels of noise (y axis) and with different output ranges (x axis) is indicated by the color bar. The black dashed line in (a) is a schematic representation of noise (i.e., the standard deviation of the response distributions, green and blue). The square bracket above the response distributions in (a) indicates the output range (i.e., the maximal difference of the mean values of the response distributions). Increasing the output range (b) and reducing the noise level (c) decrease the overlap (uncertainty) between response distributions observed in (a). (d) Acquired information is largest (maximal) when noise is minimized and the output range is maximized.

The overlap between response distributions can be reduced in two ways (11). First, the response distributions can be made more distinct by separating their means while preserving their dispersion (e.g., their standard deviation). This implies that the range of outputs produced by the stimuli will increase (Fig 1b). The second way to decrease the overlap between the response distributions is to decrease their dispersion, while keeping the mean constant. This is equivalent to reducing noise, which allows detecting the signal with increasing precision (Fig 1c). Thus, information acquisition is maximized when the output range is maximized and noise is minimized (Fig 1d).

Various mechanisms can either reduce noise or increase the output range to improve information acquisition (10,11,15–17). These include feedback loops and protein-protein interactions that reduce the level of noise (12,18,19), and increasing the number of molecules, which increase the output range (20,21). As a result, we know that noise – and thus also information acquisition – can be tuned within some limits (20,22–26). However, it is less clear how signaling pathways adjust their kinetic parameters, such as the association and dissociation rate of binding molecules, to minimize noise or increase the output range to respond efficiently to an environmental signal.

While a few studies have explored the effect of kinetic parameters on information transmission in signaling pathways, small gene networks, and gene expression systems (10,11,16,26), most of these studies did not explicitly model the molecular interactions involved in signaling. Therefore, they provide little intuition about why and how the kinetic properties of molecular processes affect information acquisition. To overcome this limitation, we use models that explicitly include all relevant molecular interactions and that do not make any a priori assumptions about statistical properties of noise. With these models, we analyze how the kinetic properties of the reversible binding interactions used by signaling pathways affect the relationship between noise, output range and information acquisition. First, we study the relationship between noise, output range and information in the reversible binding of two molecules that represent a signal and a receptor. We then analyze how information is transmitted in a chain of consecutive binding interactions. We then focus on information acquisition in gene regulation by the reversible binding of a TF to the DNA. Finally, we assemble all these components to help us understand information acquisition in a simple model of a linear signaling pathway. Our results show that, contrary to what is expected, under a broad range of biochemically sensible parameters, a noisy dynamic of the molecular interactions increases information acquisition in signaling pathways.

Results

The models

We study multiple models that represent either different fundamental steps of a signaling pathway or a complete pathway. All these models include an input or signal molecule S and an output O that conveys information about the signal's value. We

quantified (1) noise as the average standard deviation of the response distributions, (2) the output range as the maximal difference of the means of the response distributions, and (3) information as the mutual information between the signal and the output (see Methods). We estimate these quantities through at least 1000 stochastic simulations for each of n evenly distributed values of the number of signal molecules (N_s) within the interval $[N_{Smax}/n, N_{Smax}]$.

In all our models, the signal is detected by reversibly binding of a molecule to either a receptor R or to a DNA binding site (DNA_{bs}). Hence, all models contain at least one reversible binding interaction between molecules. We describe the affinity of two reversibly binding molecules with the equilibrium constant $K_{eq}(M)=k_d/k_a$, where k_d and k_a represent the dissociation and association rate, respectively. The equilibrium constant represents the concentration of free signal molecules at which half of the receptors are bound to a signal molecule. As the equilibrium constant decreases, the concentration of signal molecules required to occupy 50% of the receptors decreases too. Hence, smaller K_{eq} means higher affinity.

Throughout this paper, we will refer to low, intermediate and high affinities in the following sense. A low affinity refers to an equilibrium constant that is much higher than the maximal concentration of the signal. A high affinity refers to an equilibrium constant that is much lower than the maximal concentration of the signal. Finally, an intermediate affinity refers to an equilibrium constant that is between the minimal and the maximal concentration of the signal. In all our models we considered biologically meaningful values of all biochemical parameters (See Methods and S1-4 Tables).

Reversible binding of molecules

We first study the reversible binding between two types of molecules, S and R that form RS complexes (Fig 2a). In this highly simplified model of an information transmission system, we consider the number of RS complexes as the output or response that conveys information about the presence of the signal S . Although this notation is suggestive of interactions between a signal (S) and a receptor (R), our framework below applies to any other reversible binding of two molecules that form a complex. However, for simplicity, we will refer to R molecules as receptors, and to S molecules as signal molecules.

Fig 2. Noise, output range and information in the reversible binding of molecules. (a) Schematic representation of reversible binding involving a receptor and a signal as examples. k_a and k_d correspond to the association and dissociation rate, respectively. (b) Acquired information, output range, and noise for receptor-ligand binding at different affinity values (K_{eq}). The red circle denotes the affinity at which mutual information between signal and output is optimized. Information, noise, and output range are normalized by their respective maximal values. Further panels show the system's behavior at (c,d) low affinity ($K_{eq}=10^{-5}$), (e,f) high affinity ($K_{eq}=10^{-9}$), and (g,h) intermediate affinity ($K_{eq}=10^{-7}$;g-h). (c, e, g) show the temporal dynamic of the receptor-signal complexes (N_{RS}) at three different concentrations of the signal S . (d, f, h) show response distributions at these signal concentrations (see color legend at the bottom of the figure).

We asked how noise, output range and information change with the affinity between receptor and signal molecules. For this analysis, we assumed that the concentration of the receptors is 10^{-8}M and that the concentration of signal molecules lies within the range $[10^{-8}\text{M}, 10^{-6}\text{M}]$. This means that the maximal number of signal molecules is greater than the total number of receptors. Our simulations allow us to distinguish three regimes as a function of affinity. First, when affinity is low, noise, output range and information are close to zero (Fig 2b). The reason is that few receptor-signal complexes

form at low affinity, regardless of the signal concentration (S1 Fig). Consequently, the number of receptor-signal complexes (N_{RS}) is close to zero for all values of the signal concentration (i.e., $N_{RS} \rightarrow 0$ for all N_S as $K_{eq} \rightarrow \infty$; Fig 2c). For this reason, response distributions overlap greatly (Fig 2d), causing the output range and information to approach zero (Fig 2b). In addition, noise is also close to zero because the number of receptor-signal complexes fluctuates little (Fig 2c).

Second, when affinity is high, receptors are saturated at most or all signal concentrations (S1 Fig), such that the number of receptor-signal complexes is equal to the total number of receptors (i.e., $N_{RS} \rightarrow N_R$ for all N_S as $K_{eq} \rightarrow 0$; Fig 2e), and the output range approaches zero (Fig 2b). Noise approaches zero as well (Fig 2b), because the number of receptor-signal complexes barely fluctuates from its large value (Fig 2e), and because the overlap between response distributions is large (Fig 2f), acquired information approaches zero as well (Fig. 2b).

All this changes at intermediate affinities, where receptors can acquire information about the number of signal molecules, because receptors are no longer mainly saturated or unoccupied. Instead, the number of receptor-signal complexes fluctuates (Fig 2g). These fluctuations increase noise, but at the same time they permit that the mean number of receptor-signal complexes differs for different number of signal molecules. As a result, the output range increases (Fig 2b), which decreases the overlap between output distributions (Fig 2h), increasing the acquired information (Fig 2b). These observations show that a noisy signal-receptor binding dynamics can be beneficial when a receptor is to acquire information about a signal. Remarkably, the amount of acquired information is maximal when noise is close to its maximally possible value (Fig 2b). In

sum, if information is acquired through reversible binding interactions, the binding kinetics that yield close to maximal noise also yield close to maximal information.

Next, we wondered if one can preserve the low noise of high and low affinity binding, while increasing the output range to maximize information acquisition. In doing so, we studied how changes in signal and receptor concentrations affect information, noise and output range. These concentrations, together with the affinity, completely determine the system's behavior. We varied these concentrations in two different ways. First, we varied the concentrations of the receptors and signal molecules by identical amounts, which keep the ratio of receptors to signal molecules constant. Second, we only varied the concentration of the signal, which changes this ratio. In both cases, we found the same qualitative relationship between noise, output range, and information as before, as long as the maximal number of signal molecules is in excess of the number of receptors. In other words, efficient information acquisition requires high levels of noise and a high output range (Fig 3a-c; S2 Fig). The higher the signal and receptor concentrations are, the lower are the affinity values required for efficient information acquisition (Fig 3a). The reason is that a receptor's affinity to its signal needs to decrease as signal concentration increases; otherwise receptors become saturated and no longer detect signal changes effectively.

Fig 3. Acquired information, output range and noise at different signal concentrations. Contour plots of (a) information, (b) output range, and (c) noise at different receptor signal ratios N_{RT}/N_{Smax} (x axis), and at different affinities (y axis). (a-c) The large red-dashed rectangles circumscribe biologically common ranges of receptor-signal affinities ($[10^{-6}M, 10^{-9}M]$) and N_{RT}/N_{Smax} ratios ($N_{RT}=10^{-8}$ and $10^{-7}M \leq N_{Smax} \leq 10^{-5}M$). The small red-dashed rectangle circumscribes the region where $N_{Smax}=N_{RT}$ and where the system is noise-free, reaches the maximally possible output range, and where information acquisition is 'perfect'. Acquired information, output range and noise are plotted from minimally to maximally observed values, color-coded as indicated by the color bar.

The only scenario where low noise allows maximal information acquisition requires fewer signal molecules than receptors (Fig 3a-c, lower right corners small red rectangle). As an extreme case, one can think of a system with an infinitely large number of receptors, a finite number of signal molecules, and extremely high receptor-signal affinity. In such a system all signal molecules are bound to receptors. Because there are fewer signal molecules than receptors, the system effectively ‘counts’ the number of signal molecules through the number of receptor-signal complexes. Notice that experimentally measured affinity values between receptors and signals, are not extremely high. Instead, they are of the same order of magnitude as signal and receptor concentrations (11,28). In our simulations, these are the affinity values where high information acquisition entails high noise (Fig 3a-c big red rectangle), suggesting that biological system operate in the noisy regime. In sum, under biologically feasible conditions, high noise is necessary for (maximal) information acquisition.

Consecutive reversible binding interactions

In a signaling pathway, the binding of a signal to a receptor is usually the first of a chain of reversible events. These events include the reversible modification of one or more intermediary signaling molecules, and they usually terminate in the reversible binding of transcriptional regulators, such as a transcription factors (TF), to DNA. TF-DNA binding differs from other signaling binding interactions because regulated genes have one or few copies in any one genome, and any one regulated gene harbors few – usually fewer than ten – TF-binding sites (29,30). In the simplest signaling pathways, signal-bound receptors can directly regulate transcription without intervening signaling steps (31).

To study how TF-DNA binding might affect information acquisition in such a pathway, we model two consecutive reversible binding interactions. They represent the formation of a receptor-signal complex, and the binding of this complex to a DNA binding site (DNA_{bs}). We assumed that the concentration of receptor molecules is $10^{-8}M$, that the concentration of signal molecules lies in the interval $[10^{-8}M, 10^{-6}M]$ and that a single DNA binding site mediates transcriptional regulation. The receptor-signal- DNA_{bs} complex represents the ultimate output of the system that harbors information about the signal.

We analyzed how the affinities of both the receptor to the signal ($K_{eqR,S}$) and of the receptor-signal complex to DNA ($K_{eqRS,D}$) affect the acquisition of information, output range and noise. As in the simpler two-molecule system, the receptor is able to detect different signal concentration at intermediary affinity values between the receptor and the signal, where the largest output ranges, which are necessary for the receptor to sense the signal, are produced with high levels of noise (Fig 4a-c).

Fig 4. Information, and noise in a pair of reversible binding interactions. Contour plots of (a and d) noise, output range (b and e) and information acquisition (c and f) in the receptor-signal complex (RS ; a-c) and in the receptor-signal- DNA_{bs} complex (RSD ; d-f) as a function of the affinities between both the receptor and the signal ($K_{eqR,S}$), and the receptor-signal complex with the downstream molecule ($K_{eqRS,D}$). Red-dashed rectangles circumscribe biologically sensible receptor-signal DNA affinities ($[10^{-8}M, 10^{-13}M]$) and receptor signal affinities ($[10^{-6}M, 10^{-9}M]$). White-dashed rectangles delineate the region of maximal information acquisition at the receptor-signal- DNA_{bs} level. Acquired information, output range and noise are plotted from minimally to maximally observed values, color-coded as indicated by the color bar.

To subsequently transmit the information acquired by the receptor-signal complex to the receptor-signal- DNA_{bs} complex, DNA binding needs to be subject to the same kind of fluctuations. Such fluctuations only occur at intermediary affinity values between the receptor-signal complex and the DNA, otherwise the DNA binding site is either almost always bound (saturated) or unbound by the receptor-signal molecules. The fluctuations in receptor-signal-DNA binding increase noise (Fig 4d), but they also lead to different probabilities of DNA binding for different concentrations of the receptor-signal complex, which increases the output range (Fig 4e). As a result, the acquisition of information increases (Fig 4f). In sum, information about a signal is obtained at intermediate values of both affinities (compare the white rectangles, indicating the region with maximal information at the receptor-signal- DNA_{bs} level in Fig 4). We also note that the affinities leading to high information acquisition and high noise in our model are similar to experimentally measured affinities between receptors and signals, as well as between transcriptional regulators to DNA (Fig 4, large red rectangles). Repeating our analysis with up to ten DNA binding sites leads to the same conclusions (S3 Fig): A noisy dynamic is essential to acquire information.

Gene expression system

At the end of signaling pathways stands the regulation of gene expression, which usually requires reversible binding (of a transcription factor to DNA), and additionally involves the synthesis and degradation of mRNA and protein. To find out whether the observed relationship between information, output range, and noise is similar in the presence of synthesis and degradation, we modeled the regulation of gene expression mediated by a transcription factor that reversibly binds to DNA. We assumed that a gene with a single DNA binding site drives transcription initiation, which occurs only when the binding site

is bound by a transcription factor. In this case, mRNA molecules are transcribed at rate k_1 , and proteins are translated from the mRNA molecules at a rate k_2 . Both mRNA and protein molecules become degraded at rates, d_1 and d_2 , respectively (Fig 5a). We considered the number of TF molecules as the signal, and the number of protein molecules N_p as the output or response.

Fig 5. Noise, output range and information in gene regulation. (a) Schematic representation of our model of gene regulation. k_a and k_d correspond to the association and dissociation rate, respectively of a TF with its DNA binding site; k_1 and k_2 correspond to the mRNA and protein synthesis rate, respectively; d_1 and d_2 correspond to the mRNA and protein degradation rates, respectively. (b) Information, output range and noise observed in numerical simulations of the system at different TF-DNA affinities (K_{eq}). Information, noise, and output range are normalized by their respective maximal values. The red circle denotes the affinity at which maximal information is acquired. System behavior at low ($K_{eq}=10^{-9}$; c and d), intermediate ($K_{eq}=10^{-11}$; e and f), and high ($K_{eq}=10^{-13}$; g and h) affinities. Temporal protein dynamics at three different TF concentrations are shown in c, e and g. Response distribution for the same simulations are shown in d, f and h. The blue dashed line in c-h marks the expected mean protein value for constitutive (unregulated, always-on) expression.

We started by analyzing how a TF's affinity to its DNA_{bs} affects the relationships between information, output range, and noise (Fig 5b). As in receptor-signal binding, at the lowest affinities, the DNA binding site is almost never bound by TF molecules, regardless of the number of TF molecules (S4a Fig). Hence, little mRNA and protein is produced, independently of the number of TF molecules (Fig 5c and S4 Fig). Response distributions are insensitive to the TF concentration, with a mean close to zero and almost no variation (Fig 5d). For this reason, both noise and output range approach zero as the affinity approaches zero, and so does the acquired information (Fig 5b).

At the highest affinities, noise shows one noticeable difference to the reversible binding of molecules (Fig 2b): it does not decrease to zero (Fig 5b). The reason is that mRNA and

protein production have ‘bursty’ dynamics with large excursions from a base line. This bursty dynamics comes from the stochastic nature of mRNA and protein production, which causes fluctuations in the concentration of both kinds of molecules (23). For this reason, gene expression is intrinsically noisy. In particular, for a gene with constitutive expression, the expected number of protein molecules N_p and its expected noise (standard deviation) are equal to $E(N_p) = (k_1/d_2)(k_2/d_1)$ and $E(\sigma(N_p)) = \sqrt{(k_1/d_2)(k_2/d_1)^2}$, respectively (24). At the highest affinities, the system behaves like a constitutive gene, because TF molecules are almost always bound to the DNA_{bs} (S4a Fig), and the regulated gene is thus constantly transcribed. Accordingly, in our simulations, the mean number of expressed protein molecules and its standard deviation are close to the expected values for a constitutive gene independently of the number of TF molecules (Fig 5e; S4c Fig; S5 Fig). Because the mean is similar for all number of TF molecules, the output range tends to zero (Fig 5b). However, the amount of noise is higher than zero (Fig 5b) and similar to that expected for a constitutive gene (S5 Fig). As a result, all response distributions are similar to those for the highest concentration of the transcription factor (Fig 5f), and the amount of acquired information about this concentration is small (Fig 5b).

At intermediary affinities, the DNA binding site is not always bound by a TF. It fluctuates between a bound (active) state, when protein molecules are synthesized, and an unbound (inactive) state, when previously synthesized proteins are degraded (Fig 5g, S4a Fig). These fluctuations increase the noise in protein concentrations at intermediate affinities relative to low and high affinities (Fig 5b). They also increase the output range of the system (Fig. 5b). Most importantly, the probability that a binding site is bound by a TF changes with the number of TF molecules, which renders the system’s output – the

number of synthesized proteins – sensitive to its input (Fig 5g). Hence, the amount of information acquired about this input increases too (Fig 5b). In other words, noise can increase the acquisition of information also in this gene expression system. Moreover, once again, acquired information is maximal when noise is close to its maximal value (Fig 5b).

Information, noise and output range in a complete signaling pathway

In a final analysis, we assembled all of the above elements – receptor-signal binding, TF-DNA binding, and gene regulation – into a model of a simple complete signaling pathway. This pathway is akin to a nuclear hormone receptor pathway, such as the signaling pathway of estrogen, progesterone, and various other lipid-soluble signals (31). In this pathway, we quantified the amount of information about the concentration of the input (hormone) signal that is contained in the number of expressed protein molecules. This analysis confirmed our previous results. As in the simpler systems, maximal information acquisition requires high noise, which increases the pathway's sensitivity to variations in the signal (Supplementary text; S6 Fig).

Discussion

A fundamental step in signaling pathways is the reversible binding of molecules, which is necessary for the detection of a signal by receptors and for the acquisition of information about this signal. Previous experimental and theoretical work has demonstrated that biological processes, including signaling pathways and their binding interactions, are inherently noisy (23). One would thus expect that the kinetic parameters of binding interactions have evolved to minimize noise, because noise is

detrimental for the acquisition of information (13,14,20). However, we find the opposite. The kinetic parameters of signaling pathways must produce noisy binding dynamics or a signaling pathway will acquire little or no information. This is due to the nature of reversible binding interactions. Under biologically sensible parameter values and realistic concentrations of ligands and receptors, binding of molecules is noise-free only when a receptor is completely saturated with its ligand, or if it is unable to bind the ligand. In either case, information acquisition is impossible. Hence, noise in molecular binding is not just unavoidable but necessary for information acquisition in signaling pathways. Importantly, the positive role of noise for information acquisition is not limited to individual binding interactions, but also occurs in more complex systems that include gene expression regulation and more than one binding interaction.

In our models, we observe only one condition where noise is not required for information acquisition. At high signal-receptor affinity, a noise-free ‘perfect’ detection of a signal is possible when the number of receptors is greater than the number of signal molecules. However, producing more receptors than signaling molecules would incur enormous energetic costs. Relatedly, transcriptional regulation generally involves fewer than ten TF binding sites per regulated gene – the analog of a receptor in such a system (29,30) – a number that is much smaller than the average number of transcription factors per cell, which are usually in the hundreds for bacteria and in the thousands for yeast and mammal cells (32,33). Hence, a perfect detection of the number of TFs or signal molecules is not biologically plausible.

Some previous work hinted at a positive role of noise for information acquisition (11,25,27), but this work was not ideally suited to understand the mechanisms by which

noise helps increase information acquisition: It did not focus on signaling pathways, did not model molecular interactions explicitly, or it assumed that noise comes from an external source and can be made arbitrarily small. In contrast, our models represent molecular interactions explicitly, which causes noise to emerge naturally from them. In doing so, they also provide a mechanistic explanation of the relationship between noise and information acquisition. However, our models focus on the simplest molecular interactions, and they do not exhaust all possible signaling interactions. Whether other properties of signaling pathways change the way kinetic parameters affect noise and information acquisition is an important task for future work.

Our models include multiple simplifying assumptions. For example, we assumed that the numbers of signaling molecules, receptors, and transcriptional regulators are constant, whereas they may change dynamically in cells. We also considered a simple linear pathway, whereas signaling pathways usually contain regulatory motifs, such as feedback circuits and feed-forward loops (34). In addition, we did not consider molecular interactions such as dimerization (18,19). Similarly, we did not consider the costs of expressing an information processing machinery (20). Because these factors do not affect the nature of reversible binding, we suspect that they might also not reduce the positive role of noisy binding dynamics for information acquisition. However, some of them might increase information acquisition at low noise by other means. For example, some signaling mechanisms increase the amount of information acquired while decreasing noise (11,12,15–19). In contrast to the molecular interactions we study, where noise increases information acquisition by increasing a system's output range, these mechanisms maintain the output range while decreasing noise (14). It

remains to be seen how such different mechanisms interact and jointly affect how biological systems acquire information.

Methods

Reversible and consecutive molecular binding models

We consider two kinds of molecules, S (signal) and R (receptor), which can associate reversibly into receptor-signal complexes at an association rate k_a ($M^{-1}s^{-1}$), and a dissociation rate k_d (s^{-1}).

To model consecutive reversible binding steps, we assume that, first, a signal (S) and a receptor (R) reversibly associate into a receptor-signal (RS) complex. Second, this complex binds reversibly to a downstream molecule (D), such as DNA. We denote the rate of association between the signal and the receptor by k_{aRS} ($M^{-1}s^{-1}$), and that of dissociation by k_{dRS} (s^{-1}). Similarly, we denote the rate of association between the receptor-signal complex and the downstream molecule by $k_{aRS,D}$ ($M^{-1}s^{-1}$), and that of dissociation by k_{dRSD} (s^{-1}).

Gene expression system

We model a gene expression system where one chemical species, denoted as TF (transcription factor), binds to a DNA binding site (DNA_{bs}) to regulate the expression of a nearby gene. TF molecules associate with the DNA_{bs} at a rate k_a ($M^{-1}s^{-1}$). The dissociation of TF - DNA_{bs} complexes happens at a rate k_d (s^{-1}). In the disassociated state, no transcription occurs, and in the associated state transcription occurs at a rate k_l (s^{-1}). Transcribed mRNA molecules are degraded at a rate d_1 (s^{-1}). Finally, proteins are translated from mRNA molecules at a rate k_2 (s^{-1}), and degraded at a rate d_2 (s^{-1}).

Complete linear signaling pathway

Our model considers the reversible receptor-ligand complex formation and gene expression activation, which is mediated by the receptor-signal complex. Consequently, the parameters that govern the behavior of such a pathway are similar to those described so far, namely: 1) an association rate ($k_{a_{R,S}}$) between the signal and receptor (R) and a dissociation rate ($k_{d_{RS}}$) of the receptor-signal complexes (RS), 2) an association ($k_{a_{RS,D}}$) and a dissociation ($k_{d_{RSD}}$) rate between RS and a DNA binding site (DNA_{bs}), and 3) a rate of gene transcription (mRNA synthesis, k_1), mRNA degradation (d_1), protein synthesis (k_2), and protein degradation (d_2).

Stochastic simulations

We simulated the behavior of the models described above using Gillespie's discrete stochastic simulation algorithm (35), using the numpy python package for scientific computing (<http://www.numpy.org/>). Gillespie's algorithm captures the stochastic nature of chemical systems. It assumes a well-stirred and thermally equilibrated system with constant volume and temperature. The algorithm requires the probability p_j that a chemical reaction R_j occurs in a given time interval $[t, t+\tau)$. Any such probability p_j is proportional to both the reaction rate and the number of reacting molecules. Notice that for first-order reactions, such as the dissociation of a molecular complex into its constituent molecules, p_j is independent of the volume in which the reaction takes place. In contrast, p_j is inversely proportional to the volume in second-order reactions, such as the association of two molecules. For the reversible molecular binding modeled here, the probabilities p_a and p_d that two molecules associate and dissociate, respectively, are proportional to

$$p_a = \frac{k_a}{VN_A} N_a N_b$$

$$p_d = k_d N_c$$

where V is the reaction volume, N_A is Avogadro's number, and N_a , N_b , and N_c are the numbers of molecules of the two chemical species a and b and of the complex c .

The probabilities p_{mR_s} , p_{mR_d} , p_{P_s} and p_{P_d} of a mRNA transcription, mRNA degradation, protein synthesis, and protein degradation event are given by

$$p_{mR_s} = k_1 N_c$$

$$p_{mR_d} = d_1 N_{mR}$$

$$p_{P_s} = k_2 N_{mR}$$

$$p_{P_d} = d_2 N_P,$$

respectively. In these expressions, the quantities N_{mR} and N_P are the numbers of mRNA molecules and of protein molecules, respectively. We model a haploid organism with only a single DNA binding site, corresponding to a single regulated gene. In this case, the probability of mRNA synthesis can be reduced to

$$p_{mR_s} = k_1$$

when the binding site is bound by transcription factor ($N_c=1$) and to

$$p_{mR_s} = 0$$

when the binding site is unbound ($N_c=0$).

Initial conditions for simulations

To determine the initial conditions of the system, we first calculated the expected number of complexes formed as

$$\hat{N}_C = N_{B_T} \frac{N_A}{K_{eq} + N_A}$$

The quantity N_A is the number of signal or TF molecules and N_{B_T} is the total number of receptor molecules or DNA binding sites. Notice that $\hat{N}_C$ is a real number, and for the specific case of the TF-DNA_{bs} interaction it can only take a value between 0 and 1. Thus, $\hat{N}_{RS}$ equals to the probability that the DNA binding site is bound by a transcription factor. However, for the receptor signal complexes, it represents the number of complexes formed. We selected the initial state of the number of complexes (N_{Ci}) at random with binomial probability for the binding site. However, we selected the closest integer to $\hat{N}_C$ for the receptor signal case. Then we defined

$$N_{Ai} = N_A - N_{Ci}$$

$$N_{Bi} = N_{B_T} - N_{Ci}$$

as the initial state of the number of free signal or TF molecules (N_{Ai}) and of the receptors or binding sites (N_{Bi}). Finally, as the initial state of the number of mRNA and protein molecules we used

$$N_{mRi} = \hat{N}_C \frac{k_1}{d_1}$$

$$N_{Pi} = \hat{N}_C \frac{k_1 k_2}{d_1 d_2}$$

which is the expected average number of mRNA and protein molecules for a constitutively expressed gene (23), multiplied by the probability that the DNA binding site is bound by a TF molecule.

Information quantification

The number of molecules of any chemical species in a cell or in a unit volume fluctuates, because molecules are produced and decay stochastically, and because they undergo

random Brownian motion caused by thermal vibrations. We use Shannon's entropy to quantify the unpredictability caused by such stochastic fluctuations in a signal as

$$H(\Pr(S)) = - \sum_{N_S = \frac{N_{Smax}}{n}}^{N_{Smax}} p(N_S) \log_2 p(N_S),$$

where $\Pr(S)$ is the probability distribution of the signal, and $p(N_S)$ is the probability that the system contains N molecules of the signal. In our models the signal represents either a molecular signal or cue (2).

For all our analyses, we performed at least 1000 simulations for each of n different numbers of signal molecules that were evenly distributed within the interval $[N_{Smax}/n, N_{Smax}]$ ($n \leq N_{Smax}$). For this reason

$$H(\Pr(S)) = \log_2 n$$

Signals trigger changes in a cell's state that produce a response or output (O) of the system, such as the production of molecules. A cell acquires information when the output O reflects (fully or partially) the value of S . This information can be quantified via the mutual information:

$$I(S;O) = H(\Pr(S)) - H(\Pr(S|O)),$$

which is equal to the entropy $H(\Pr(S))$ minus the conditional entropy $H(\Pr(S|O))$, which represents the entropy of S given that O is known (13). In other words, the mutual information I quantifies the acquired information as the amount of information that an output chemical species O harbors about a signal S .

Noise quantification and output range

The systems we model produce a probabilistic response for any given quantity N_S of the signal. This response can thus be represented as a conditional probability distribution:

$$\Pr(0 < N_O < N_{O_{max}} | S = N_S),$$

where N_O and $N_{O_{max}}$ are the number and maximal number of output molecules, respectively. We estimated this response distribution through at least 1000 replicate simulations of the system for each value of N_S . We then quantified noise as the standard deviation of the response distributions, averaged over all n possible values of N_S :

$$\bar{\sigma} = \frac{1}{n} \sum_{N_S = N_{Smin}}^{N_{Smax}} \sigma(\Pr(O|S = N_S)),$$

We define the output range as the difference between the maximal and minimal mean value of all response distributions.

Parameter values

Our simulations considered the following biologically sensible parameter ranges. The association and dissociation constants k_a and k_d of reversible complex formation define the equilibrium constant $K_{eq} = k_d/k_a$ (M), which we used in our simulations. The smaller K_{eq} becomes, the more association becomes favored over dissociation (28). In particular, for the binding between ligands and (nuclear) receptors, we used values of K_{eq} within the interval $[10^{-6}\text{M}, 10^{-9}\text{M}]$, because the micromolar to nanomolar range is common for such complexes (28,32,36–38). For TF-DNA binding, empirical data suggests that usually $K_{eq} < 10^{-8}$ and can reach picomolar (10^{-12}M) or even smaller values (28,32,36–38). Thus, we used values in the interval $[10^{-8}\text{M}, 10^{-12}\text{M}]$.

For mRNA, experimentally measured half-lives usually lie in the range of seconds to hours (39–42). Protein half-lives typically lie between hours and days (41,43). Taking all

this information into consideration, we chose mRNA half-lives within the interval [1min,30min], and protein half-lives where within [15min, 3h]. We assume that the ratio k_2/k_1 , which describes the speed of the protein synthesis rate relative to the mRNA synthesis rate, exceeds 1.0 (44). Because the residence time of transcription factors on DNA lies within seconds to hours (45,46), we assumed a residence time within this interval [10sec,2h]

Finally, we always considered concentrations of molecules to lie within the interval [10^{-9} M, 10^{-6} M], because these are typical concentration of most molecules within a cell or a nucleus (47). Notice that for some of our simulation we also needed to explore values outside these ranges. Specific parameter values used for each simulation are listed in supplementary tables 1-4.

Acknowledgments

We acknowledge support from the European Research Council under Grant Agreement No. 739874, by Swiss National Science Foundation grant 31003A_172887, as well as by the University Priority Research Program in Evolutionary Biology.

References

1. Bowsher CG, Swain PS. Environmental sensing, information transfer, and cellular decision-making. *Current Opinion in Biotechnology*. 2014 Aug;28:149–55.
2. Laidre ME, Johnstone RA. Animal signals. *Current Biology*. 2013 Sep;23(18):R829–33.
3. Boyer PD, Krebs EG, Tamanoi F, editors. *The Enzymes*. 3d ed. New York: Academic Press; 1970. 1 p.
4. Cantley LC. *Signal transduction: principles, pathways, and processes*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory; 2014.
5. Tsimring LS. Noise in biology. *Reports on Progress in Physics*. 2014 Feb 1;77(2):026601.
6. Pilpel Y. Noise in Biological Systems: Pros, Cons, and Mechanisms of Control. In:

- Castrillo JI, Oliver SG, editors. *Yeast Systems Biology* [Internet]. Totowa, NJ: Humana Press; 2011 [cited 2019 Aug 9]. p. 407–25. Available from: http://link.springer.com/10.1007/978-1-61779-173-4_23
7. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, et al. Accurate information transmission through dynamic biochemical signaling networks. *Science*. 2014 Dec 12;346(6215):1370–3.
8. Uda S, Saito TH, Kudo T, Kokaji T, Tsuchiya T, Kubota H, et al. Robustness and Compensation of Information Transmission of Signaling Pathways. *Science*. 2013 Aug 2;341(6145):558–61.
9. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science*. 2011 Oct 21;334(6054):354–8.
10. Tkacik G, Callan CG, Bialek W. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences*. 2008 Aug 26;105(34):12265–70.
11. Tkačik G, Walczak AM, Bialek W. Optimizing information flow in small genetic networks. *Physical Review E* [Internet]. 2009 Sep 29 [cited 2019 Aug 9];80(3). Available from: <https://link.aps.org/doi/10.1103/PhysRevE.80.031920>
12. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*. 2001 Jul 17;98(15):8614–9.
13. Cover TM, Thomas JA. *Elements of information theory*. 2nd ed. Hoboken, N.J: Wiley-Interscience; 2006. 748 p.
14. Rhee A, Cheong R, Levchenko A. The application of information theory to biochemical signaling systems. *Physical Biology*. 2012 Aug 1;9(4):045011.
15. Rieckh G, Tkačik G. Noise and Information Transmission in Promoters with Multiple Internal States. *Biophysical Journal*. 2014 Mar;106(5):1194–204.
16. Tkačik G, Walczak AM, Bialek W. Optimizing information flow in small genetic networks. III. A self-interacting gene. *Physical Review E* [Internet]. 2012 Apr 5 [cited 2019 Aug 9];85(4). Available from: <https://link.aps.org/doi/10.1103/PhysRevE.85.041903>
17. Walczak AM, Tkačik G, Bialek W. Optimizing information flow in small genetic networks. II. Feed-forward interactions. *Physical Review E* [Internet]. 2010 Apr 6 [cited 2019 Aug 9];81(4). Available from: <https://link.aps.org/doi/10.1103/PhysRevE.81.041905>
18. Liu F-Y, Lo S-C, Shu C-C. The Reaction of Dimerization by Itself Reduces the Noise Intensity of the Protein Monomer. *Scientific Reports* [Internet]. 2019 Dec [cited 2019 Aug 9];9(1). Available from: <http://www.nature.com/articles/s41598-019-39611-6>
19. Ghim C-M, Almaas E. Genetic noise control via protein oligomerization. *BMC Systems Biology* [Internet]. 2008 Dec [cited 2019 Aug 9];2(1). Available from: <https://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-2-94>
20. ten Wolde PR, Becker NB, Ouldridge TE, Mugler A. Fundamental Limits to Cellular Sensing. *Journal of Statistical Physics*. 2016 Mar;162(5):1395–424.
21. Pahle J, Green AK, Dixon CJ, Kummer U. Information transfer in signaling pathways: A study using coupled simulated and experimental data. *BMC Bioinformatics*. 2008 Dec;9(1):139.
22. Richard M, Yvert G. How does evolution tune biological noise? *Frontiers in Genetics* [Internet]. 2014 Oct 28 [cited 2019 Aug 9];5. Available from: <http://journal.frontiersin.org/article/10.3389/fgene.2014.00374/abstract>
23. Raj A, van Oudenaarden A. Single-Molecule Approaches to Stochastic Gene

- Expression. Annual Review of Biophysics. 2009 Jun;38(1):255–70.
24. Kepler TB, Elston TC. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. Biophysical Journal. 2001 Dec;81(6):3116–36.
25. Lestas I, Vinnicombe G, Paulsson J. Fundamental limits on the suppression of molecular fluctuations. Nature. 2010 Sep;467(7312):174–8.
26. Bialek W, Setayeshgar S. Physical limits to biochemical signaling. Proc Natl Acad Sci U S A. 2005 Jul 19;102(29):10040.
27. Bowsher CG, Voliotis M, Swain PS. The Fidelity of Dynamic Signaling by Noisy Biomolecular Networks. Haugh JM, editor. PLoS Comput Biol. 2013 Mar 28;9(3):e1002965.
28. Corzo J. Time, the forgotten dimension of ligand binding teaching. Biochemistry and Molecular Biology Education. 2006 Nov;34(6):413–6.
29. Lusk RW, Eisen MB. Evolutionary Mirages: Selection on Binding Site Composition Creates the Illusion of Conserved Grammars in Drosophila Enhancers. Barsh GS, editor. PLoS Genetics. 2010 Jan 22;6(1):e1000829.
30. Bilu Y, Barkai N. The design of transcription-factor binding sites is affected by combinatorial regulation. Genome Biology. 2005 Dec 2;6(12):R103.
31. Sever R, Glass CK. Signaling by Nuclear Receptors. Cold Spring Harbor Perspectives in Biology. 2013 Mar 1;5(3):a016709–a016709.
32. Biggin MD. Animal Transcription Networks as Highly Connected, Quantitative Continua. Developmental Cell. 2011 Oct;21(4):611–26.
33. Ho B, Baryshnikova A, Brown GW. Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Systems. 2018 Feb;6(2):192–205.e3.
34. Alon U. Network motifs: theory and experimental approaches. Nature Reviews Genetics. 2007 Jun 1;8:450.
35. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977 Dec;81(25):2340–61.
36. Belikov S, Berg OG, Wrangé Ö. Quantification of transcription factor-DNA binding affinity in a living cell. Nucleic Acids Research. 2015 Dec 10;44(7):3045–58.
37. Fisher WW, Li JJ, Hammonds AS, Brown JB, Pfeiffer BD, Weizmann R, et al. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in Drosophila. Proceedings of the National Academy of Sciences. 2012 Dec 26;109(52):21330–5.
38. Li X-Y, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biology. 2011;12(4):R34.
39. Lugowski A, Nicholson B, Rissland OS. Determining mRNA half-lives on a transcriptome-wide scale. Methods. 2018 Mar;137:90–8.
40. Laalami S, Zig L, Putzer H. Initiation of mRNA decay in bacteria. Cellular and Molecular Life Sciences. 2014 May;71(10):1799–828.
41. Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature. 2011 May;473(7347):337–42.
42. Sharova LV, Sharov AA, Nedorezov T, Piao Y, Shaik N, Ko MSH. Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells. DNA Research. 2009 Jan 8;16(1):45–58.

43. Toyama BH, Hetzer MW. Protein homeostasis: live long, won't prosper. *Nature Reviews Molecular Cell Biology*. 2013 Jan;14(1):55–61.
44. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nature Communications* [Internet]. 2019 Dec [cited 2019 Aug 9];10(1). Available from: <http://www.nature.com/articles/s41467-018-07391-8>
45. Morisaki T, Müller WG, Golob N, Mazza D, McNally JG. Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nature Communications* [Internet]. 2014 Dec [cited 2019 Aug 9];5(1). Available from: <http://www.nature.com/articles/ncomms5456>
46. Mueller F, Stasevich TJ, Mazza D, McNally JG. Quantifying transcription factor kinetics: At work or at play? *Critical Reviews in Biochemistry and Molecular Biology*. 2013 Sep;48(5):492–514.
47. Legewie S, Herzel H, Westerhoff HV, Blüthgen N. Recurrent design patterns in the feedback regulation of the mammalian signalling network. *Molecular Systems Biology*. 2008 Jan;4(1):190.

Supporting information captions

S1 Fig. Mean number of receptor-signal complexes ($\bar{N}_{RS}$) formed at different affinity values (K_{eq}). The maximum number of receptor-signal complexes for this simulation is 50 (see S1 Table).

S2 Fig. Noise, output range and information observed in numerical simulations of the receptor-signal system at different affinities (K_{eq}) and with different concentrations of R and S . a) $R=10^{-9}M$ and $S=[10^{-9}M, 10^{-7}M]$. b) $R=10^{-8}M$ and $S=[10^{-8}M, 10^{-6}M]$. c) $R=10^{-7}M$ and $S=[10^{-7}M, 10^{-5}M]$. Information, noise, and output range are normalized by their respective maximal values.

S3 Fig. Information and noise in two consecutive binding interactions with ten DNA_{bs}. Contour plots of (a and d) noise, output range (b and e) and information acquisition (c and f) in the receptor-signal complex (RS ; a-c) and in the receptor-signal-DNA_{bs} complex (RSD ; d-f) as a function of the affinities between both the receptor and the signal ($K_{eqR,S}$), and the receptor-signal complex with the downstream molecule ($K_{eqRS,D}$). Red-dashed rectangles circumscribe biologically sensible receptor-signal DNA affinities ($[10^{-8}M, 10^{-13}M]$) and receptor signal affinities ($[10^{-6}M, 10^{-9}M]$). White-dashed rectangles delineate the region of maximal information acquisition at the receptor-signal-DNA_{bs} level. Acquired information, output range and noise are plotted from minimally to maximally observed values, color-coded as indicated by the color bar.

S4 Fig. (a) Probability of TF-DNA binding, (b) mean number of mRNA molecules ($\bar{N}_{mRNA}$), and (c) mean number of protein molecules ($\bar{N}_P$), as a function of TF-DNA affinity (K_{eq}). For the parameters we used (S3 Table), the expected

mean number of mRNA and protein molecules produced from a constitutively expressed gene is 2 and 100, respectively.

S5 Fig. Non-normalized values of noise in the number of proteins. The black line is the standard deviation expected for the parameters used in this simulation (S3 Table).

S6 Fig. Contour plot of acquired information and noise in a complete signaling pathway. (a) Information and (b) noise for different affinities between receptor and signal molecules ($K_{eqR,S}$, y axis), and receptor-signal complexes with a DNA binding site ($K_{eqRS,D}$, x axis) in a model of a simple lineal signaling pathway where a signal bound receptor can bind DNA and regulate gene expression. The number of signal molecules constitute the input into this pathway, and the number of protein molecules expressed from the regulated gene constitute the output. The red-dashed rectangle show experimentally measured affinity values between receptors and signals ($[10^{-6}\text{M}, 10^{-9}\text{M}]$) and between transcriptional regulators and DNA ($[10^{-9}\text{M}, 10^{-13}\text{M}]$). The amount of acquired information and noise level are indicated in the color bar.

Low S concentration (10^{-8}M) Intermediate S concentration (10^{-7}M) High S concentration (10^{-6}M)

a**Information** **b****Output range** **c****Noise**

bioRxiv preprint doi: <https://doi.org/10.1101/762989>; this version posted September 9, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.