Supporting Information

Bowsher and Swain 10.1073/pnas.1119407109

SI Text

A Brief Primer on Conditional Expectations

Conditional expectations are not commonly used outside of probability and statistics. We present here a short introduction and list of their properties.

For two random variables X and Y, the conditional expectation $E[X|Y]$ is itself a random variable because it is a function of the random variable Y. For continuous random variables, $E[X|Y]$ is defined as

$$E[X|Y=y] = \int dx\, x\, p(x|y), \qquad \text{[S1]}$$

where $p(x|y)$ is the conditional probability density of X given Y. It satisfies $p(x|y) = p(x,y)/p(y)$, by Bayes's rule. The conditional expectation has the following properties for any three random variables X, Y and Z:

i. If X and Y are independent, then

$$E[X|Y] = E[X]. \qquad \text{[S2]}$$

ii. If X and Y are conditionally independent given Z, then

$$E[X|Y,Z] = E[X|Z]. \qquad \text{[S3]}$$

iii. For real constants a and b

$$E[aX+bY|Z] = aE[X|Z] + bE[Y|Z]. \qquad \text{[S4]}$$

iv. If knowing the random variable Z implies that X is known, then

$$E[XY|Z] = XE[Y|Z] \qquad \text{[S5]}$$

and so $E[g(Z)|Z] = g(Z)$ for any (measurable) function $g(Z)$.

v. If knowing Z implies that X is known, then

$$E[Y|X] = E[E[Y|Z]|X]. \qquad \text{[S6]}$$

vi. For any Y,

$$E[E[X|Y]] = E[X]. \qquad \text{[S7]}$$

The conditional expectation $E[X|Y]$ is, in the sense of minimizing the mean squared error, the best approximation to X. For any real-valued function $f(Y)$, it can be shown that

$$E[(X-f(Y))^2] \geq E[(X-E[X|Y])^2]. \qquad \text{[S8]}$$

Useful Additional Properties of Conditionally Independent Reporters

Suppose that Z_t' is conditionally independent of Z_t given $(X, \mathscr{H}_t)$, where X is a time invariant random variable (for example, one controlled in an experiment or set by the environment) and $\mathscr{H}_t$ is some history. Then

$$\begin{aligned}\mathrm{Cov}[Z_t, Z_t'|X] &= E\{E[Z_tZ_t'|X,\mathscr{H}_t]|X\} - E\{E[Z_t|X,\mathscr{H}_t]|X\} \\ &\quad \cdot E\{E[Z_t'|X,\mathscr{H}_t]|X\} \\ &= \mathrm{Cov}[E[Z_t|X,\mathscr{H}_t], E[Z_t'|X,\mathscr{H}_t]|X],\end{aligned} \qquad \text{[S9]}$$

where we have used the conditional independence of Z_t and Z_t'. If Z_t' is also first-moment conjugate to Z_t for $(X, \mathscr{H}_t)$, then we shall show that the covariance of the reporters conditional on X identifies the second component of the decomposition of conditional variance, $V[Z_t|X] = E[V[Z_t|X,\mathscr{H}_t]|X] + V[E[Z_t|X,\mathscr{H}_t]|X]$. When the two reporters are first-moment conjugate, $E[Z_t|X,\mathscr{H}_t] = E[Z_t'|X,\mathscr{H}_t]$, which implies that $E[Z_t|X] = E[Z_t'|X]$, and therefore

$$\begin{aligned}\mathrm{Cov}[Z_t, Z_t'|X] &= \mathrm{Cov}[E[Z_t|X,\mathscr{H}_t], E[Z_t'|X,\mathscr{H}_t]|X] \\ &= V\{E[Z_t|X,\mathscr{H}_t]|X\}.\end{aligned} \qquad \text{[S10]}$$

Consequently, the average conditional covariance (averaging with respect to the distribution of X) gives $E\{\mathrm{Cov}[Z_t, Z_t'|X]\} = E\{V[E[Z_t|X,\mathscr{H}_t]|X]\}$.

Transcriptional and Translational Variance: Reaction Network and Parameter Values Used in Simulations for Fig. 2*B*

We used the *Facile* compiler (1) and the *EasyStoch* simulator (2), which encodes the Gibson–Bruck (3) version of the Gillespie algorithm (4). We specify the model and the parameters used to generate the data underlying Fig. 2*B* in the main paper in the format employed by *Facile* (see Table S1). Comments are marked with a hash and the initial numbers of molecules are denoted with N. Any chemical species not specified initially has zero molecules.

For convenience, we simultaneously simulate three reporters (the original system of interest, a copy, and a bicistronic reporter). To model extrinsic fluctuations in v_0 the rate of transcription, we use 'dummy' chemical species, S1, S2, and S3, to control the propensity of transcription. Only one of these species exists at any given time and transitions between the three forms of S generate transitions in the value of v_0. We denote the protein reporter equivalent to Z in Fig. 2*A* as B, the reporter equivalent to Z' as A, and the reporter equivalent to Z'' as C.

Interpreting Scatter Plots of Measurements of Reporters

Plotting single-cell measurements of one reporter, Z, against measurements for a reporter conjugate to Z given some history $Y^{\mathscr{H}}$ (denoted Z') gives a scatter plot where the extents of the scatter of points parallel and perpendicular to the $Z = Z'$ diagonal measure different components of the variance. A typical example is shown in Fig. S1. Each point represents measurements of a reporter and its conjugate in a single cell and has coordinates (Z, Z'). Note that each reporter has the same mean value from the conditions of conjugacy. For each point, we can define $d_\perp$, which measures the distance from the point to the $Z = Z'$ diagonal, and $d_\parallel$, which measures the distance along the diagonal that the point lies from the point corresponding to the mean value (Fig. S1).

We can show that the mean value of $d_\perp^2$, the spread of the points perpendicular to the $Z = Z'$ diagonal, satisfies

$$E[d_\perp^2] = \frac{1}{2}E\left[(Z - Z')^2\right] \quad \textbf{[S11]}$$

because the point of intersection (red dot in Fig. S1) is $((Z + Z')/2, (Z + Z')/2)$. For any point (Z, Z'), $d_\perp^2$ is then

$$d_\perp^2 = \left(Z' - \frac{Z + Z'}{2}\right)^2 + \left(Z - \frac{Z + Z'}{2}\right)^2, \quad \textbf{[S12]}$$

giving Eq. **S11** taking expectations. The right-hand side of Eq. **S11** corresponds generally to a sum of terms in the decomposition of variance, with the particular sum being determined by the choice of the conditioning used to select the conjugate reporter.

Similarly, the mean value of $d_\parallel^2$, the spread of the points along the diagonal, satisfies

$$E[d_\parallel^2] = \mathrm{Cov}[Z, Z'] + \frac{1}{2}E\left[(Z - Z')^2\right] + \mathrm{Cov}[Z, Z']. \quad \textbf{[S13]}$$

If $E[(Z - Z')^2]/2$ corresponds to a sum of terms in the decomposition of variance, then $\mathrm{Cov}[Z, Z']$ corresponds to the sum of the remaining terms. By definition,

$$\begin{aligned} d_\parallel^2 &= \left(\frac{Z + Z'}{2} - E[Z]\right)^2 + \left(\frac{Z + Z'}{2} - E[Z]\right)^2 \\ &= \frac{1}{2}\left(Z - E[Z] + Z' - E[Z]\right)^2, \end{aligned} \quad \textbf{[S14]}$$

implying that

$$E[d_\parallel^2] = V[Z] + \mathrm{Cov}[Z, Z'] \quad \textbf{[S15]}$$

and giving Eq. **S13** because $V[Z] = E[(Z - Z')^2]/2 + \mathrm{Cov}[Z, Z']$ (Eq. 17 in the main text).

Translational Variation: Analyzing the Data of Kollmann et al.

Using similar arguments to those given in the *Appendix* section of the main text, we can also show that

$$\begin{aligned} E[(Z - Z_c)^2] = E\{V[Z|Y^{\mathscr{H}}]\} + E\{V[Z_c|Y^{\mathscr{H}}]\} \\ + E\{(E[Z|Y^{\mathscr{H}}] - E[Z_c|Y^{\mathscr{H}}])^2\} \end{aligned} \quad \textbf{[S16]}$$

if Z and Z_c are conditionally independent given some history $Y^{\mathscr{H}}$. Consequently,

$$E[(Z - Z_c)^2] \geq E\{V[Z|Y^{\mathscr{H}}]\} + E\{V[Z_c|Y^{\mathscr{H}}]\}, \quad \textbf{[S17]}$$

where both terms on the right-hand side can be measured using conjugate reporters with equal second conditional moments: $E\{V[Z|Y^{\mathscr{H}}]\}$ is equal to $E[(Z - Z')^2]/2$ if Z' is a reporter conjugate to Z given $Y^{\mathscr{H}}$, and $E\{V[Z_c|Y^{\mathscr{H}}]\}$ is equal to $E[(Z_c - Z_c')^2]/2$ if Z_c' is a reporter conjugate to Z_c given $Y^{\mathscr{H}}$.

Kollmann et al. (5) measured gene expression of the chemotaxis proteins CheY, tagged with YFP, and CheZ, tagged with CFP, with both proteins expressed from the same mRNA. CheY-YFP and CheZ-CFP should be conditionally independent given the joint history of the levels of the bicistronic mRNA, M, and the stochastic variables extrinsic to gene expression, Y_e. We can therefore use Eq. **S17** with CheY-YFP denoted by Z, CheZ-CFP denoted by Z_c, and $Y^{\mathscr{H}}$ being the joint history of M and Y_e. Kollmann et al. found that

$$E[(Z - Z_c)^2] \simeq 2 \times 0.2^2 E[Z]E[Z_c], \quad \textbf{[S18]}$$

where Z and Z_c are measured in fluorescence units (5). Consequently,

$$\begin{aligned} &\frac{1}{2}[\overbrace{E\{V[Z|(M, Y_e)^{\mathscr{H}}]\}}^{\text{translational for CheY}} + \overbrace{E\{V[Z_c|(M, Y_e)^{\mathscr{H}}]\}}^{\text{translational for CheZ}}] \\ &\quad \leq 0.2^2 E[Z]E[Z_c], \end{aligned} \quad \textbf{[S19]}$$

and therefore the average translational variance for the two proteins, normalized by the product of their mean fluorescences, is less than 0.2^2.

Finding Bounds on Components of the Variance Decomposition

Suppose that the reporter Z_t'' is conditionally independent of the reporter Z_t given $(X, \mathscr{H}_t)$, where X is again a time invariant random variable. To find a lower bound on $V\{E[Z|X, \mathscr{H}_t]\}$, we begin with a conditional form of the Cauchy–Schwarz inequality:

$$\mathrm{Cov}[W, W''|X]^2 \leq V[W|X] \cdot V[W''|X], \quad \textbf{[S20]}$$

for arbitrary random variables W and W''. From Eq. **S9**,

$$\mathrm{Cov}[Z_t, Z_t''|X] = \mathrm{Cov}[E[Z_t|X, \mathscr{H}_t], E[Z_t''|X, \mathscr{H}_t]|X],$$

and therefore the Cauchy–Schwarz inequality directly implies that

$$V\{E[Z_t|X, \mathscr{H}_t]|X\} \geq \frac{\mathrm{Cov}[Z_t, Z_t''|X]^2}{V\left\{E[Z_t''|X, \mathscr{H}_t]\middle|X\right\}}, \quad \textbf{[S21]}$$

where the denominator $V\{E[Z_t''|X, \mathscr{H}_t]|X\}$ can itself be measured by the covariance (conditional on X) between Z_t'' and a reporter conjugate to Z_t'' for the conditioning $(X, \mathscr{H}_t)$. The lower bound in Eq. **S21** becomes an equality when $E[Z_t|X, \mathscr{H}_t]$ is a linear function of $E[Z_t''|X, \mathscr{H}_t]$.

Distinguishing Variation due to Gene Expression from Variation due to Signal Transduction: Analyzing the Data of Colman-Lerner et al.

Colman-Lerner et al. used the promoter for PRM1 driving YFP to quantify the response of budding yeast to pheromone (6). From Eqs. **7** and **8**, we can write an inequality for the variation generated by signal transduction:

$$\begin{aligned} &\overbrace{E\{V[E[Z|(Y_{e\setminus T}, T)^{\mathscr{H}}, X]|Y^{\mathscr{H}}_{e\setminus T}, X]|X\}}^{\text{from signal transduction given } X} \\ &\quad \leq \mathrm{Cov}[Z, Z'|X] - \frac{\mathrm{Cov}[Z, Z_c|X]^2}{\mathrm{Cov}[Z_c, Z_c'|X]}, \end{aligned} \quad \textbf{[S22]}$$

where Z is a reporter for the output of the system; Z' is a reporter conjugate to Z given the history of all extrinsic variables; Z_c is a reporter for a constitutively expressed gene; and Z_c' is a reporter conjugate to Z_c given the history of extrinsic variables (Fig. 2*C*). Alejandro Colman-Lerner kindly provided: average fluorescence measurements (the total fluorescence in individual cells divided by the area of the cells) of a strain expressing both YFP and CFP from two copies of the promoter for PRM1 across a population of

172 cells (corresponding to Z and Z' in Eq. **S22**); average fluorescence measurements of a strain expressing both YFP and CFP from two copies of the promoter for ACT1 (actin) across a population of 292 cells (corresponding to Z_c and Z_c' in Eq. **S22**); and average fluorescence measurements of a strain expressing CFP from the promoter of ACT1 and YFP from the promoter of PRM1 across a population of 233 cells.

To adjust for the different brightness of CFP and YFP, we corrected the measurements of CFP in the strain expressing CFP and YFP from the promoter of PRM1 to have the same median as the measurements of YFP. We corrected the measurements of CFP in the strain expressing CFP and YFP from the promoter of ACT1 and the measurements of CFP in the strain expressing CFP from ACT1 and YFP from the promoter of PRM1 to have the same median as measurements of YFP in the strain expressing CFP and YFP from the promoter for ACT1.

Both the CFP and YFP measurements should also be corrected for cellular autofluorescence, although autofluorescence is less of a problem for YFP because it is brighter. We were unable to correct the data for autofluorescence, and numerical values should be interpreted with this caveat.

The Informational Fraction of Variance

Some Intuition. For an output Z and input X, let $E[Z|X] = \mu(X)$ and $V[Z|X] = \sigma^2(X)$. Then,

$$V[Z] = V\{E[Z|X]\} + E\{V[Z|X]\} = V[\mu(X)] + E[\sigma^2(X)], \quad \textbf{[S23]}$$

and the informational fraction of the output variance is

$$\iota_{Z|X} = \frac{V\{E[Z|X]\}}{V[Z]} = \frac{V[\mu(X)]}{V[\mu(X)] + E[\sigma^2(X)]} = \left(1 + \frac{E[\sigma^2(X)]}{V[\mu(X)]}\right)^{-1}. \quad \textbf{[S24]}$$

In the main text, we denote $\iota_{Z|X}$ by ι_Z.

Imagine drawing two independent realizations of the input X from its distribution, denoted by X_1, X_2. Write the corresponding expected outputs conditional on the realized inputs as μ_1 and μ_2, where $\mu_i = \mu(X_i)$. Then, the typical distance between the two conditional means obtained is

$$\frac{1}{2}E[(\mu_1 - \mu_2)^2] = V[\mu_i], \quad \textbf{[S25]}$$

because $E[\mu_1\mu_2] = E[\mu_1]^2$. The expected conditional variance for each draw is simply $E[\sigma_i^2]$. Therefore, as the informational fraction tends to one,

$$\frac{1}{2}E[(\mu_1 - \mu_2)^2] \gg E[\sigma^2(X)], \quad \textbf{[S26]}$$

and the typical distance between the means of a pair of conditional distributions for the output Z becomes much larger than the expected variability or "width" of those distributions: The conditional output distributions typically overlap less. Heuristically, each output distribution is less likely to overlap with another and the system should become more efficient at transmitting information. We make these ideas more precise by providing formal connections between the informational fraction and information theory below.

Input and Output with a Jointly Gaussian Distribution and a General Upper Bound on the Conditional Entropy of the Output. Consider the input and output (X, Z) to be a continuous random vector, with the support of Z equal to $(-\infty, \infty)$. Let z be the rescaled output with variance equal to 1, $z = Z/V[Z]^{1/2}$. The rescaling affects neither the informational fraction, nor the mutual information of input and output (7). Note that $1 - \iota_{z|X} = E\{V[z|X]\}$ and that the entropy of a Gaussian distribution with variance v is equal to $\frac{1}{2}\ln(2\pi e v)$. Now $V[z|X = x] \geq \frac{1}{2\pi e}\exp\{2h(z|X = x)\}$ because the Gaussian distribution has the maximum entropy for a given variance. It follows after taking the expectation of both sides of the inequality and applying Jensen's inequality that

$$1 - \iota_{Z|X} = E\{V[z|X]\} \geq \frac{1}{2\pi e}\exp\{2h(z|X)\},$$

where $h(z|X) = E_x[h(z|X = x)]$. Therefore, an upper bound for the conditional entropy of the rescaled output is given by

$$h(z|X) \leq \frac{1}{2}\ln\{2\pi e[1 - \iota_{Z|X}]\}. \quad \textbf{[S27]}$$

The upper bound decreases as $\iota_{Z|X}$ increases, placing an upper limit on how uncertain the output can be given the state of the input.

When the signaling mechanism obeys "Gaussian statistics," or more precisely the conditional distribution of output given input, $p(z|X)$, is Gaussian with variance not depending on X, it is seen that $V[z|X]\} = \frac{1}{2\pi e}\exp\{2h(z|X)\}$ and therefore Eq. **S27** holds with equality in this case. If the input X is also normally distributed then (X, Z) has a bivariate normal distribution and z is therefore normally distributed with variance equal to 1. The mutual information $I(X; Z) = I(X; z)$. It follows directly that

$$I(X; Z) = h(z) - h(z|X) = \frac{1}{2}\ln\{2\pi e\} - \frac{1}{2}\ln\{2\pi e[1 - \iota_{Z|X}]\} = -\frac{1}{2}\ln\{1 - \iota_{Z|X}\}, \quad \textbf{[S28]}$$

which is familiar on recognizing that $\iota_{Z|X} = \mathrm{Corr}[Z, X]^2$, because $E[Z|X]$ is a linear function of X (see Eq. **S31**) for a bivariate normal distribution. Throughout, we define the correlation of any two random variables T, U to be $\mathrm{Corr}[T, U] = \mathrm{Cov}[T, U]/V[T]^{\frac{1}{2}}V[U]^{\frac{1}{2}}$.

A Lower Bound on Information Capacity Is Set by $\iota_{Z|X}$. We will now prove that when the conditional means $\mu(X)$ are different for all values of X, the information capacity C of the biochemical mechanism satisfies the lower bound given by

$$C = \sup_{p(X)} I(X; Z) \geq \frac{1}{2}\ln[1 - \iota_{Z|\tilde{X}}]^{-1}, \quad \textbf{[S29]}$$

where the supremum ("maximum") is taken over a set of possible input distributions, $\mathcal{S}$. The informational fraction is evaluated for an input distribution corresponding to a Gaussian distribution for $\mu(X)$. The higher the informational fraction, the larger the lower bound on the capacity.

Consider the mechanism

$$W \xrightarrow{f} X \overset{p(Z|X)}{\rightarrow} Z, \quad \textbf{[S30]}$$

where $p(Z|X)$ represents the biochemical transduction, $Z \in (-\infty, \infty)$, $W \in (-\infty, \infty)$, and the function f is a continuously differentiable, one-to-one mapping. Where necessary, we transform the biophysical output (which is often positive), for example by taking its logarithm, so that Z is real-valued. For any distribution $p(W)$, we have that $I(W; Z) = I(X; Z)$ (7).

We will need the following result. For any two random variables T, U, the informational fraction satisfies the equality

$$\iota_{U|T} = \mathrm{Corr}(U, E[U|T])^2, \quad \textbf{[S31]}$$

from the definition of correlation and because

$$\begin{aligned}\mathrm{Cov}(U, E[U|T]) &= E\{E[(U - E[U])(E[U|T] - E[U])|T]\} \\ &= E\{E[U|T]^2 - 2E[U]E[U|T] + E[U]^2\} \\ &= V\{E[U|T]\}.\end{aligned}$$

Therefore, $\iota_{Z|W} = \mathrm{Corr}(Z, E[Z|W])^2$. Furthermore, if $E[Z|W]$ is a linear (affine) function of W, then $\mathrm{Corr}(Z, E[Z|W]) = \mathrm{Corr}(Z, W)$ and $\iota_{Z|W} = \mathrm{Corr}(Z, W)^2$ for any distribution $p(W)$.

Notice that for the mechanism in Eq. **S30**, the random variables $E[Z|X]$ and $E[Z|W]$ are equal because f is an invertible function and conditioning on X is therefore equivalent to conditioning on W (mathematically, the conditioning sigma field $\sigma(X) = \sigma(W)$). It follows immediately that the corresponding informational fractions are equal:

$$\iota_{Z|W} = \iota_{Z|X}, \quad \textbf{[S32]}$$

where we have used $V[Z|X] = V[Z|W]$.

The essential insights in the proof of the lower bound in Eq. **S29** are to approach the problem via the "augmented" mechanism in Eq. **S30** and to then notice that a certain choice of the function f will result in $E[Z|W]$ being a linear function of W. This choice is useful because it is known how to bound the mutual information from below when the input is Gaussian, using the squared correlation of input and output. As we have seen, when the conditional expectation is a linear function of W, $\iota_{Z|W} = \mathrm{Corr}(Z, W)^2$. The choice of f is to set

$$X = \mu^{-1}(W) \quad \text{where } \mu(x) = E[Z|X = x].$$

The inverse biochemical "response" function, μ^{-1}, is expected to be smooth (continuously differentiable) for biophysically reasonable response functions $\mu(x)$. When f is set equal to μ^{-1} in Eq. **S30**,

$$E[Z|W = w] = E[Z|X = \mu^{-1}(w)] = \mu(\mu^{-1}(w)) = w,$$

or more concisely $E[Z|W] = E[Z|X] = W$, which is the linearity in W we set out to achieve.

It now follows, recalling Eq. **S32**, that $\mathrm{Corr}[Z, W]^2 = \iota_{Z|X}$ for any distribution $p(W)$ and the implied input distribution $p(X)$. Let $\tilde{W}$ denote the artificial input when that random variable has a Gaussian distribution. Then, $I(\tilde{W}; Z) \geq \frac{1}{2}\ln[1 - \mathrm{Corr}[Z, \tilde{W}]^2]^{-1}$ (8). We have now shown that

$$\begin{aligned}C \geq I(\tilde{X}; Z) = I(\tilde{W}; Z) &\geq \frac{1}{2}\ln[1 - \mathrm{Corr}[Z, \tilde{W}]^2]^{-1} \\ &= \frac{1}{2}\ln[1 - \iota_{Z|\tilde{X}}]^{-1},\end{aligned}$$

where $\tilde{X} = \mu^{-1}(\tilde{W})$ has the distribution implied by the Gaussian distribution of $\tilde{W}$.

In biology, because natural input distributions have not been widely measured, the set $\mathcal{S}$ of possible input distributions $p(X)$ must be specified by the investigator, and a range of choices for $\mathcal{S}$ may be entertained. To implement the capacity bound in Eq. **S29**, one can range over choices for the mean and variance of the Gaussian $\tilde{W}$, excluding those choices that imply a distribution $p(X)$ that one wants to omit from $\mathcal{S}$. Armed with the function $\sigma^2(X) = V[Z|X]$, both $p(\tilde{X})$ and the informational fraction $\iota_{Z|\tilde{X}}$ can be computed by Monte Carlo sampling from $p(\tilde{W})$, using the relation $\tilde{X} = \mu^{-1}(\tilde{W})$. To implement the capacity bound, the other input distributions in $\mathcal{S}$ need not be specified. Finally, the informational fraction $\iota_{Z|\tilde{X}}$ should be maximized over the set of distributions $p(\tilde{X})$ given by the allowed means and variances for $\tilde{W}$.

As a simple illustration, consider the Gaussian noise channel of information theory given by $Z = gX + \eta_{Z|X}$, where g is a constant and $\eta_{Z|X}$ is normally distributed with zero mean and a constant variance $\sigma^2_{Z|X}$ that is not dependent on X. Let $\mathcal{S}$ consist of input distributions satisfying $E[X] = 0$ and $V[X] \leq \sigma^2$. Because $W = gX$ here, we set $E[\tilde{W}] = 0$ and $V[\tilde{W}] \leq g^2\sigma^2$. The informational fraction $\iota_{Z|\tilde{X}} = g^2V[\tilde{X}]/(g^2V[\tilde{X}] + \sigma^2_{Z|X})$, which is maximized by setting $V[\tilde{X}] = \sigma^2$, which implies $V[\tilde{W}] = g^2\sigma^2$. The corresponding, maximized lower bound for the capacity given by Eq. **S29** is then equal to $\frac{1}{2}\ln[1 - \iota_{Z|\tilde{X}}]^{-1} = \frac{1}{2}\ln[1 + (g^2\sigma^2/\sigma^2_{Z|X})]$, which is exactly equal to the capacity of the Gaussian noise channel with input "power" constraint σ^2. Our lower bound on the information capacity is a tight one for the Gaussian channel.

Information Transfer When $\iota_{Z|X}$ Is Large. Consider now a setting in which the biochemical mechanism and the input distribution can vary as $n \to \infty$, where n labels the sequence of mechanisms and input distributions. Suppose that $\iota_{Z_n|X_n}$, the informational fraction for Z_n, converges to its maximum value of 1 and that the unconditional distribution $p(z_n)$ of the rescaled output of the signaling mechanism does not become ever less uncertain as $n \to \infty$. By uncertainty in the continuous case (or differential entropy), we mean the logarithm of the effective volume of the smallest set that contains most of the probability (9). We will give a concrete biochemical example of such asymptotic behavior below.

More precisely, suppose $\iota_{Z_n|X_n} \to 1$ and that $h(z_n) \not\to -\infty$ (or, equivalently, $h(z_n)$ is bounded below by a constant for all n), where $z_n = Z_n/V[Z_n]^{\frac{1}{2}}$ as before. Then, Eq. **S27** implies that $-h(z_n|X_n) \to +\infty$ and hence

$$I(X_n; Z_n) = I(X_n; z_n) \to \infty \quad \text{as } \iota_{Z_n|X_n} \to 1. \quad \textbf{[S33]}$$

Biophysically reasonable transduction mechanisms are expected to give rise to unconditional distributions for the rescaled output, $p(z_n)$, that reflect the uncertainty of the input rather than having differential entropy that is unbounded below. If the input distribution varies as the limit is taken, we assume its uncertainty (differential entropy) does not become ever less as $n \to \infty$.

As an example of such asymptotics, suppose we hold the input distribution constant for simplicity and consider the linear noise approximation (LNA) of output at time t, for which (10)

$$Z_{t,n} = \Omega_n\phi(t, X) + \Omega_n^{1/2}\xi(t, X), \quad \textbf{[S34]}$$

where Ω_n is the system size, $\phi(t, X)$ is the deterministic solution for output concentration at time t, and the random variable $\xi(t, X)$ can be shown in the case of the LNA not to depend on Ω_n (11). Notice that Eq. **S34** makes no assumption of Gaussianity. Let $\tilde{Z}_{t,n} \coloneqq Z_{t,n}/\Omega_n$ denote the output concentration. We can see that, as the system size $\Omega_n \to \infty$, then $\iota_{Z_{t,n}|X} \to 1$ because

$$\frac{E\{V[\tilde{Z}_{t,n}|X]\}}{V\{E[\tilde{Z}_{t,n}|X]\}} = \frac{E\{V[\xi(t, X)|X]\}}{\Omega_n V[\phi(t, X)] + \Omega_n^{1/2}\mathrm{Cov}\{\phi(t, X), E[\xi(t, X)|X]\} + V\{E[\xi(t, X)|X]\}} = \mathcal{O}(\Omega_n^{-1}).$$

Furthermore, it follows from Eq. **S34** that $\tilde{Z}_{t,n} \to \phi(t, X)$: The output concentration converges (almost surely) to the deterministic solution, which is a function of X. Let $\tilde{z}_{t,n} = \tilde{Z}_{t,n} V[\tilde{Z}_{t,n}]^{-1/2}$ be the rescaled output with a variance of 1. The differential entropy of the rescaled output $h(\tilde{z}_{t,n}) \to h(\phi(t, X)) - \frac{1}{2} \ln V\{\phi(t, X)\}$, under suitable regularity conditions. We make the mild assumption that the distribution of the continuous input is such that $|h(\phi(t, X))| < \infty$ and $V\{\phi(t, X)\} < \infty$. It then follows from Eq. **S27** and the above argument that $I(X; Z_{t,n}) = I(X; \tilde{z}_{t,n}) \to \infty$ as $\Omega_n \to \infty$. Information transfer becomes perfect in the limit of large system size. Because the LNA tells us about moments but not distributions, it is not clear how to prove this property without making use of the informational fraction and the implied properties when the informational fraction tends to its maximal value of 1.

Determining the Informational Fraction for Osmosensing in Budding Yeast

Pelet et al. (12) used YFP to report gene expression from the STL1 promoter for six different concentrations of extracellular salt. They recorded fluorescence levels from approximately 1,000 cells for each concentration of salt.

Letting P_i denoting the probability of the environment having a salt concentration equal to S_i, then the informational fraction for a given probability distribution of extracellular salt is

$$\text{informational fraction} = \frac{V\{E[Z|S]\}}{V[Z]} = \frac{\sum_i P_i E[Z|S_i]^2 - (\sum_i P_i E[Z|S_i])^2}{\sum_i P_i E[Z^2|S_i] - (\sum_i P_i E[Z|S_i])^2}, \quad \textbf{[S35]}$$

where we have used $V[W] = E[W^2] - E[W]^2$. Therefore, if $y_{i,j}$ is the average fluorescence level of YFP in the jth cell (the total fluorescence in that cell divided by the area of the cell) when the salt concentration is S_i and there are N_i such cells, then our empirical measure of the informational fraction is given by

$$\frac{\sum_i P_i \left(\frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}\right)^2 - \left[\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}\right]^2}{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}^2 - \left[\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}\right]^2}. \quad \textbf{[S36]}$$

We exhaustively searched the possible probability distributions over the six different concentrations of extracellular salt and determined the probability distributions that have high informational fractions. We discretized P_i (to have 21 equally spaced values, between and including the values zero and one) and looped through all possible values of P_i for each i, calculating the informational fraction only when $\sum_i P_i = 1$.

For two reporters that are conjugate given the history of the stochastic variables extrinsic to gene expression, then the total extrinsic fraction for a particular concentration of salt is defined as the ratio of the covariance of the two reporters to the variance of the output Z:

$$\text{total extrinsic fraction} = \frac{\text{Cov}[Z, Z'|S_i]}{V[Z]} = \frac{\sum_i P_i E[ZZ'|S_i] - (\sum_i P_i E[Z|S_i])(\sum_i P_i E[Z'|S_i])}{\sum_i P_i E[Z^2|S_i] - (\sum_i P_i E[Z|S_i])^2} \quad \textbf{[S37]}$$

for the experiments of Pelet *et al.* If $c_{i,j}$ is the average fluorescence measured from the CFP reporter in the jth cell when the concentration of salt is S_i, then our empirical measure of the total extrinsic fraction is given by

$$\frac{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j} c_{i,j} - \left(\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}\right)\left(\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} c_{i,j}\right)}{\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}^2 - \left[\sum_i P_i \frac{1}{N_i} \sum_{j=1}^{N_i} y_{i,j}\right]^2}. \quad \textbf{[S38]}$$

The two fluorescent proteins, CFP and YFP, have different brightness, and we multiply each $c_{i,j}$ by a correction factor so that the median of the YFP measurements is equal to the median of the CFP measurements for each concentration of salt.

Both the CFP and YFP measurements should also be corrected for cellular autofluorescence, although autofluorescence is less of a problem for YFP because it is brighter. We were unable to correct the data for autofluorescence, and numerical values should be interpreted with this caveat.

Calculation of the Variance Components from the Chemical Master Equation

The conjugate reporter method allows us to find analytical expressions for the components of the variance. Consider gene expression with one extrinsic variable (Fig. S2*A*). In a mathematical model, this extrinsic variable corresponds to a fluctuating propensity for a particular reaction (2, 13), and we will begin with a fluctuating propensity for transcription. If this propensity has three states reflecting, for example, three states of the extracellular environment (Fig. S2*B*), then we can define κ_{01} to be the probability per unit time of transitioning from the state 0 to state 1 (conditional on being in state 0); κ_{10} to be the probability per unit time of transitioning back; κ_{12} to be the probability per unit time of transitioning from state 1 to state 2; and κ_{21} to be the probability per unit time of transitioning back. With all the κ_{ij} identical, we used this model to generate the data for Fig. 2*B*.

Initially, we will consider the decomposition of the variance into intrinsic and extrinsic components,

$$V[Z(t)] = E\{V[Z(t)|v_{0,t}^{\mathscr{H}}]\} + V\{E[Z(t)|v_{0,t}^{\mathscr{H}}]\}, \quad \textbf{[S39]}$$

and therefore require reporters that are conditionally independent given the history of v_0, the propensity for the transcriptional reaction, and that have the same means conditional on $v_{0,t}^{\mathscr{H}}$. An identical copy of the system exposed to the same fluctuations in v_0 satisfies both these conditions (Fig. S2*A*). We will denote the number of mRNAs from each copy of the system as m_1 and m_2 and the number of proteins from each copy as n_1 and n_2. The probability of having m_1 mRNAs and n_1 proteins from the first copy and m_2 mRNAs and n_2 proteins from the second is $P(m_1, n_1, m_2, n_2, v_0^{(i)}, t)$, with i denoting the state of the extrinsic variable. For brevity, we will write $P^{(i)}$ for $P(m_1, n_1, m_2, n_2, v_0^{(i)}, t)$ and only explicitly write (with subscripts) the number of molecules when these differ from either m_1, n_1, m_2, or n_2. The corresponding master equations for the dual reporter systems are then (see Fig. S2*A* for definitions of the parameters)

$$\frac{\partial P^{(i)}}{\partial t} = v_0^{(i)}[P^{(i)}_{m_1-1} - P^{(i)}] + d_0[(m_1+1)P^{(i)}_{m_1+1} - m_1 P^{(i)}] + d_1[(n_1+1)P^{(i)}_{n_1+1} - n_1 P^{(i)}] + v_1 m_1[P^{(i)}_{n_1-1} - P^{(i)}] + v_0^{(i)}[P^{(i)}_{m_2-1} - P^{(i)}] + d_0[(m_2+1)P^{(i)}_{m_2+1} - m_2 P^{(i)}] + d_1[(n_2+1)P^{(i)}_{n_2+1} - n_2 P^{(i)}] + v_1 m_2[P^{(i)}_{n_2-1} - P^{(i)}] + \begin{cases} \kappa_{10}P^{(1)} - \kappa_{01}P^{(0)} & \text{if } i=0 \\ \kappa_{01}P^{(0)} - (\kappa_{10}+\kappa_{12})P^{(1)} + \kappa_{21}P^{(2)} & \text{if } i=1 \\ \kappa_{12}P^{(1)} - \kappa_{21}P^{(2)} & \text{if } i=2 \end{cases}, \quad \textbf{[S40]}$$

where there is one equation for each state of the extrinsic variable (here, $v_0^{(i)}$).

We can solve Eq. **S40** exactly for the moments of the probability distribution $P^{(i)}$. We will use $\mathbf{s}$ to represent the vector of numbers of species, $\mathbf{s} = [m_1, n_1, m_2, n_2]$, and, for brevity, we will use angled brackets to denote expectations:

$$\langle f(\mathbf{s})\rangle_i = \sum_{\mathbf{s}} P(\mathbf{s}, v_0^{(i)}) f(\mathbf{s}) = \sum_{\mathbf{s}} P^{(i)} f(\mathbf{s}) \quad \textbf{[S41]}$$

for any function $f(\mathbf{s})$ and where the expectation is taken with the particular value of v_0 fixed. By multiplying Eq. **S40** by either m_1 or m_2 and summing over all states described by $P^{(i)}$ (over all values of m_1, m_2, n_1, and n_2), we find that the mean mRNA for either copy then obeys

$$\frac{\partial \langle m\rangle_i}{\partial t} = v_0^{(i)} \sum_{\mathbf{s}} P^{(i)} - d_0\langle m\rangle_i + \begin{cases} \kappa_{10}\langle m\rangle_1 - \kappa_{01}\langle m\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle m\rangle_0 - (\kappa_{10}+\kappa_{12})\langle m\rangle_1 + \kappa_{21}\langle m\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle m\rangle_1 - \kappa_{21}\langle m\rangle_2 & \text{if } i=2 \end{cases}, \quad \textbf{[S42]}$$

where the sum over $P^{(i)}$ is over all states of the system for a given v_0. Performing this summation in Eq. **S40** gives

$$\frac{\partial}{\partial t}\sum_{\mathbf{s}} P^{(0)} = \kappa_{10}\sum_{\mathbf{s}} P^{(1)} - \kappa_{01}\sum_{\mathbf{s}} P^{(0)},$$
$$\frac{\partial}{\partial t}\sum_{\mathbf{s}} P^{(1)} = \kappa_{01}\sum_{\mathbf{s}} P^{(0)} - (\kappa_{10}+\kappa_{12})\sum_{\mathbf{s}} P^{(1)} + \kappa_{21}\sum_{\mathbf{s}} P^{(2)},$$
$$\frac{\partial}{\partial t}\sum_{\mathbf{s}} P^{(2)} = \kappa_{12}\sum_{\mathbf{s}} P^{(1)} - \kappa_{21}\sum_{\mathbf{s}} P^{(2)}, \quad \textbf{[S43]}$$

and so

$$\sum_{\mathbf{s}} P^{(0)} = \frac{\kappa_{10}\kappa_{21}}{\kappa_{10}\kappa_{21} + \kappa_{01}\kappa_{12} + \kappa_{01}\kappa_{21}},$$
$$\sum_{\mathbf{s}} P^{(1)} = \frac{\kappa_{01}\kappa_{21}}{\kappa_{10}\kappa_{21} + \kappa_{01}\kappa_{12} + \kappa_{01}\kappa_{21}},$$
$$\sum_{\mathbf{s}} P^{(2)} = \frac{\kappa_{01}\kappa_{12}}{\kappa_{10}\kappa_{21} + \kappa_{01}\kappa_{12} + \kappa_{01}\kappa_{21}} \quad \textbf{[S44]}$$

at steady state. The mean protein for either copy satisfies

$$\frac{\partial \langle n\rangle_i}{\partial t} = v_1\langle m\rangle_i - d_1\langle n\rangle_i + \begin{cases} \kappa_{10}\langle n\rangle_1 - \kappa_{01}\langle n\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle n\rangle_0 - (\kappa_{10}+\kappa_{12})\langle n\rangle_1 + \kappa_{21}\langle n\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle n\rangle_1 - \kappa_{21}\langle n\rangle_2 & \text{if } i=2 \end{cases}. \quad \textbf{[S45]}$$

The simultaneous equations, Eq. **S42** and Eq. **S45** together with Eq. **S42**, can be solved at steady state, straightforwardly with computer algebra packages such as Mathematica (Wolfram Research).

Similarly, by multiplying Eq. **S40** by, for example, m_1^2 and averaging, we can find equations for the second moments:

$$\frac{\partial \langle m^2\rangle_i}{\partial t} = 2v_0^{(i)}\langle m\rangle_i + v_0^{(i)}\sum_{\mathbf{s}} P^{(i)} + d_0\langle m\rangle_i - 2d_0\langle m^2\rangle_i + \begin{cases} \kappa_{10}\langle m^2\rangle_1 - \kappa_{01}\langle m^2\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle m^2\rangle_0 - (\kappa_{10}+\kappa_{12})\langle m^2\rangle_1 + \kappa_{21}\langle m^2\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle m^2\rangle_1 - \kappa_{21}\langle m^2\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S46]}$$

for the mean square number of molecules of mRNA;

$$\frac{\partial \langle n^2\rangle_i}{\partial t} = d_1\langle n\rangle_i + 2v_1\langle mn\rangle_i + v_1\langle m\rangle_i - 2d_1\langle n^2\rangle_i + \begin{cases} \kappa_{10}\langle n^2\rangle_1 - \kappa_{01}\langle n^2\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle n^2\rangle_0 - (\kappa_{10}+\kappa_{12})\langle n^2\rangle_1 + \kappa_{21}\langle n^2\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle n^2\rangle_1 - \kappa_{21}\langle n^2\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S47]}$$

for the mean square number of molecules of protein; and

$$\frac{\partial \langle mn\rangle_i}{\partial t} = v_0^{(i)}\langle n\rangle_i + v_1\langle m^2\rangle_i - (d_0+d_1)\langle mn\rangle_i + \begin{cases} \kappa_{10}\langle mn\rangle_1 - \kappa_{01}\langle mn\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle mn\rangle_0 - (\kappa_{10}+\kappa_{12})\langle mn\rangle_1 + \kappa_{21}\langle mn\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle mn\rangle_1 - \kappa_{21}\langle mn\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S48]}$$

for the mean product of mRNA and protein numbers. We solve Eqs. **S46**, **S47**, and **S48** at steady state simultaneously using the solutions of Eqs. **S42** and **S45** and so compute the stationary second moments.

Finally, we need the covariance between the two reporters, $\langle n_1 n_2\rangle$, to determine the extrinsic variance. From Eq. **S40**, we find three sets of coupled equations:

$$\frac{\partial \langle n_1 n_2\rangle_i}{\partial t} = 2v_1\langle m_1 n_2\rangle_i - 2d_1\langle n_1 n_2\rangle_i + \begin{cases} \kappa_{10}\langle n_1 n_2\rangle_1 - \kappa_{01}\langle n_1 n_2\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle n_1 n_2\rangle_0 - (\kappa_{10}+\kappa_{12})\langle n_1 n_2\rangle_1 + \kappa_{21}\langle n_1 n_2\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle n_1 n_2\rangle_1 - \kappa_{21}\langle n_1 n_2\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S49]}$$

to determine the covariance between the proteins;

$$\frac{\partial\langle m_1n_2\rangle_i}{\partial t} = v_0^{(i)}\langle n\rangle_i + v_1\langle m_1m_2\rangle_i - (d_0+d_1)\langle m_1n_2\rangle_i + \begin{cases} \kappa_{10}\langle m_1n_2\rangle_1 - \kappa_{01}\langle m_1n_2\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle m_1n_2\rangle_0 - (\kappa_{10}+\kappa_{12})\langle m_1n_2\rangle_1 + \kappa_{21}\langle m_1n_2\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle m_1n_2\rangle_1 - \kappa_{21}\langle m_1n_2\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S50]}$$

to determine the covariance of the mRNA of one copy of the system with the protein of another ($\langle m_1n_2\rangle = \langle m_2n_1\rangle$ from symmetry); and

$$\frac{\partial\langle m_1m_2\rangle_i}{\partial t} = 2v_0^{(i)}\langle m\rangle_i - 2d_0\langle m_1m_2\rangle_i + \begin{cases} \kappa_{10}\langle m_1m_2\rangle_1 - \kappa_{01}\langle m_1m_2\rangle_0 & \text{if } i=0 \\ \kappa_{01}\langle m_1m_2\rangle_0 - (\kappa_{10}+\kappa_{12})\langle m_1m_2\rangle_1 + \kappa_{21}\langle m_1m_2\rangle_2 & \text{if } i=1 \\ \kappa_{12}\langle m_1m_2\rangle_1 - \kappa_{21}\langle m_1m_2\rangle_2 & \text{if } i=2 \end{cases} \quad \textbf{[S51]}$$

to determine the covariance of the mRNAs from the two copies of the systems. We solve Eqs. **S49**, **S50**, and **S51** at steady state. To find the final moments, we sum the moments calculated for each state of the extrinsic variable because

$$\langle f(\mathbf{s})\rangle = \sum_{\mathbf{s},i} P(\mathbf{s}, v_0^{(i)}) f(\mathbf{s}) = \sum_{\mathbf{s},i} P^{(i)} f(\mathbf{s}) = \sum_i \langle f(\mathbf{s})\rangle_i \quad \textbf{[S52]}$$

for any function $f(\mathbf{s})$.

All these equations can be straightforwardly modified to study extrinsic fluctuations in a different kinetic rate. For example, if the translation rate fluctuates then we replace $v_0^{(i)}$ by v_0 and the translation rate v_1 by the appropriate $v_1^{(i)}$ in all the equations. To have two or more rates fluctuating (2), we can either extend the number of states of $P^{(i)}$ if the extrinsic fluctuations are uncorrelated or have more than one parameter changing with state i if the extrinsic fluctuations are correlated. Our analytical results verify the behavior found previously through simulation for particular values of parameters (2).

Extrinsic Fluctuations in Transcription Need Not Change the Form of the Intrinsic Noise. Consider extrinsic fluctuations in v_0, so that v_0 has three states: $v_0^{(i)}$, where i runs from 0 to 2. From Eq. **S44**,

$$\langle v_0\rangle = \frac{\kappa_{10}\kappa_{21}v_0^{(0)} + \kappa_{01}\kappa_{21}v_0^{(1)} + \kappa_{01}\kappa_{12}v_0^{(2)}}{\kappa_{01}\kappa_{12} + \kappa_{01}\kappa_{21} + \kappa_{10}\kappa_{21}}, \quad \textbf{[S53]}$$

and we find that

$$\langle m\rangle = \frac{\langle v_0\rangle}{d_0}; \qquad \langle n\rangle = \frac{v_1}{d_1}\langle m\rangle. \quad \textbf{[S54]}$$

To compare with previous work (14–19), we will give our results in terms of the coefficient of variation, η (the standard deviation of a variable divided by its mean). For the intrinsic noise, we have that

$$\eta_{\text{int}}^2 = \frac{1}{\langle n\rangle} + \frac{d_1}{d_0+d_1}\frac{1}{\langle m\rangle}, \quad \textbf{[S55]}$$

which has the same form for the system when no extrinsic fluctuations are present (16, 19) (Fig. S3). If we assume that $v_0^{(0)} = v_0(1-\epsilon)$, $v_0^{(1)} = v_0$, and $v_0^{(2)} = v_0(1+\epsilon)$ for a constant v_0 and ϵ and that $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$, then the extrinsic noise equals

$$\eta_{\text{ext}}^2 = \frac{d_0d_1(d_0+d_1+\kappa)}{(d_0+d_1)(d_0+\kappa)(d_1+\kappa)}\eta_{v_0}^2 \quad \textbf{[S56]}$$

and is proportional to the square of the noise in v_0, $\eta_{v_0}^2$, as expected (20). We note that $\eta_{v_0}^2 = \frac{2\epsilon^2}{3}$ with this choice of $v_0^{(i)}$.

Extrinsic Fluctuations in Translation Can Increase the Intrinsic Noise. We can proceed similarly with extrinsic fluctuations in the translation rate. If we let $v_1^{(0)} = v_1(1-\epsilon)$, $v_1^{(1)} = v_1$, and $v_1^{(2)} = v_1(1+\epsilon)$ and $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$, then

$$\langle m\rangle = \frac{v_0}{d_0}; \qquad \langle n\rangle = \frac{v_1}{d_1}\langle m\rangle \quad \textbf{[S57]}$$

and

$$\eta_{\text{int}}^2 = \frac{1}{\langle n\rangle} + \frac{d_1}{d_0+d_1}\frac{1}{\langle m\rangle}\left[1 + \frac{d_0+d_1}{d_0+d_1+\kappa}\eta_{v_1}^2\right] \quad \textbf{[S58]}$$

with

$$\eta_{\text{ext}}^2 = \frac{d_1}{d_1+\kappa}\eta_{v_1}^2. \quad \textbf{[S59]}$$

We see that the intrinsic noise is larger than the intrinsic noise of an equivalent system with no extrinsic fluctuations ($\eta_{v_1} = 0$ and Eq. **S55**) because of the factor in square brackets in Eq. **S58** (Fig. S3). This factor depends on both the magnitude and lifetime of the fluctuations in v_1, as well as the lifetime of both mRNA and protein molecules.

Extrinsic Fluctuations in the Degradation of mRNA Can Increase the Intrinsic Noise. Having extrinsic fluctuations in the degradation rates of either mRNA or protein gives more complex behaviors because such fluctuations directly affect the lifetime of fluctuations in proteins (2). Assuming, as before, that $d_0^{(0)} = d_0(1-\epsilon)$, $d_0^{(1)} = d_0$, and $d_0^{(2)} = d_0(1+\epsilon)$, that $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$, and that $\eta_{d_0}^2 < 1$, then

$$\langle m\rangle = \frac{v_0}{d_0}\left[1 + \frac{d_0}{d_0+\kappa}\eta_{d_0}^2 + \cdots\right]; \qquad \langle n\rangle = \frac{v_1}{d_1}\langle m\rangle, \quad \textbf{[S60]}$$

where we have omitted terms of order $\eta_{d_0}^4$ and higher. The intrinsic noise is approximately

$$\eta_{\text{int}}^2 \simeq \frac{1}{\langle n\rangle} + \frac{d_1}{d_0+d_1}\frac{1}{\langle m\rangle}\left[1 + \frac{d_0^2(2d_0+d_1+\kappa)}{(d_0+d_1)(d_0+\kappa)(d_0+d_1+\kappa)}\eta_{d_0}^2\right], \quad \textbf{[S61]}$$

and the extrinsic noise is approximately

$$\eta_{\text{ext}}^2 \simeq \frac{d_0d_1(d_0+d_1+\kappa)}{(d_0+d_1)(d_0+\kappa)(d_1+\kappa)}\eta_{d_0}^2, \quad \textbf{[S62]}$$

where higher order corrections in η_{d_0} have been omitted. The intrinsic noise is therefore larger than the intrinsic noise of an equivalent system with no extrinsic fluctuations ($\eta_{d_0} = 0$) providing η_{d_0} is sufficiently small (Fig. S3), and the mean number of proteins has increased above the value predicted by purely deterministic dynamics.

Extrinsic Fluctuations in the Degradation of Protein Can Decrease the Intrinsic Noise. Assuming again that $d_1^{(0)} = d_1(1-\epsilon)$, $d_1^{(1)} = d_1$, and $d_1^{(2)} = d_1(1+\epsilon)$, that $\kappa_{01} = \kappa_{10} = \kappa_{12} = \kappa_{21} = \kappa$, and that $\eta_{d_1}^2 < 1$, then

$$\langle m \rangle = \frac{v_0}{d_0}; \qquad \langle n \rangle \simeq \frac{v_1}{d_1}\langle m \rangle \left[1 + \frac{d_1}{d_1+\kappa}\eta_{d_1}^2\right], \qquad \textbf{[S63]}$$

where we ignore terms of order $\eta_{d_1}^4$ and higher. The intrinsic noise is

$$\eta_{\text{int}}^2 \simeq \frac{1}{\langle n \rangle} + \frac{d_1}{d_0+d_1}\frac{1}{\langle m \rangle} \times \left[1 - \frac{d_1^2(2d_0(d_0+d_1) - (d_1+\kappa)(2d_1+\kappa))}{(d_0+d_1)(d_1+\kappa)(d_0+d_1+\kappa)(2d_1+\kappa)}\eta_{d_1}^2\right], \qquad \textbf{[S64]}$$

and

$$\eta_{\text{ext}}^2 \simeq \frac{d_1}{d_1+\kappa}\eta_{d_1}^2, \qquad \textbf{[S65]}$$

where we have omitted higher order terms in η_{d_1}. We expect $d_0 > d_1$ (19) and $\kappa \geq d_1$ (21). The intrinsic noise is therefore typically smaller than the intrinsic noise of an equivalent system with no extrinsic fluctuations (Fig. S3), and the mean number of proteins has increased above the value predicted by deterministic dynamics.

Decomposing the Intrinsic Noise. In Eq. 3 of the main paper, we decompose the intrinsic noise into transcriptional and translational components. We further argue that a bicistronic reporter correctly measures the translational component and when combined with the original reporter for the system will allow all three components of the variance to be measured. We apply these ideas to calculate the transcriptional and translational components of the intrinsic noise when there are extrinsic fluctuations in the propensity for transcription. The master equation for a bicistronic reporter (Fig. S2*C*) is

$$\begin{aligned}\frac{\partial P^{(i)}}{\partial t} &= v_0^{(i)}[P_{m-1}^{(i)} - P^{(i)}] + d_0[(m+1)P_{m+1}^{(i)} - mP^{(i)}] \\ &+ d_1[(n_1+1)P_{n_1+1}^{(i)} - n_1P^{(i)}] + v_1 m[P_{n_1-1} - P^{(i)}] \\ &+ d_1[(n_2+1)P_{n_2+1}^{(i)} - n_2P^{(i)}] + v_1 m[P_{n_2-1} - P^{(i)}] \\ &+ \begin{cases} \kappa_{10}P^{(1)} - \kappa_{01}P^{(0)} & \text{if } i = 0 \\ \kappa_{01}P^{(0)} - (\kappa_{10}+\kappa_{12})P^{(1)} + \kappa_{21}P^{(2)} & \text{if } i = 1 \\ \kappa_{12}P^{(1)} - \kappa_{21}P^{(2)} & \text{if } i = 2 \end{cases}, \end{aligned} \qquad \textbf{[S66]}$$

where m is the number of molecules of the mRNA and we assume three different states of the extrinsic variable (here v_0). We can solve Eq. **S66** for its moments following the approach used for Eq. **S40**. The equations undergo only minor changes with some straightforward replacements (for example, $\langle m_1^{(i)} m_2^{(i)} \rangle$ becomes $\langle [m^{(i)}]^2 \rangle$). We find that the transcriptional and translational components of the intrinsic noise are

$$\eta_{\text{transc}}^2 = \frac{d_1}{d_0+d_1}\frac{1}{\langle m \rangle}; \qquad \eta_{\text{transl}}^2 = \frac{1}{\langle n \rangle}, \qquad \textbf{[S67]}$$

showing that our theoretical definitions (Eq. 3) give a natural decomposition.

Eq. **S67** implies that transcriptional variation is often greater than translational variation. Typical lifetimes of mRNA in *Escherichia coli* are several minutes, but protein numbers are often mostly reduced through dilution. Assuming a cell cycle of 50 min (22) and an average lifetime of an mRNA of 3 min (23), then $d_1/(d_0+d_1)$ is approximately 0.06, and so $\eta_{\text{transc}}^2/\eta_{\text{transl}}^2 \simeq 0.06\frac{\langle n \rangle}{\langle m \rangle}$. Consequently, transcriptional variation is greater than translational variation if $\langle n \rangle$ is approximately greater than 18 times $\langle m \rangle$, which is not uncommon: The average number of proteins per mRNA is around 540 (24).

Verifying Conditional Independences

To use conjugate reporters to determine the components of the variance of a given model, we must check that the appropriate conditional independences are satisfied. Suppose we wish to verify that two reporters Z and Z' are conditionally independent given the history $Y^{\mathscr{H}}$. Suppose further that the future dynamics of these Y variables can depend on their own histories, but (given those histories) are independent of the history of all other variables in the model. Then, informally, Z and Z' are conditionally independent given $Y^{\mathscr{H}}$ if we can first simulate the realization of Y (to time t), and then simulate two subsystems independently (or "separately") given that history of Y to obtain Z_t and Z_t'.

One of us (C.G.B.) has developed the necessary mathematical theory for verification of conditional independence properties in stochastic kinetic models (chemical master equations) in general (25, 26). An algorithm, MIDIA, that applies this theory to test conditional independences has been implemented in R and is freely available (27).

1. Siso-Nadal F, Ollivier JF, Swain PS (2007) Facile: A command-line network compiler for systems biology. *BMC Syst Biol* 1:36.
2. Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* 4:196.
3. Gibson M, Bruck J (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* 104:1876–1889.
4. Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81:2340–2361.
5. Kollmann M, Lovdok L, Bartholomé K, Timmer J, Sourjik V (2005) Design principles of a bacterial signalling network. *Nature* 438:504–507.
6. Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
7. Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. *Phys Rev E Stat Nonlin Soft Matter Phys* 69:066138.
8. Mitra PP, Stark JB (2001) Nonlinear limits to the information capacity of optical fibre communications. *Nature* 411:1027–1030.
9. Cover TM, Thomas JA (2006) *Elements of Information Theory* (Wiley-Interscience, Hoboken, NJ).
10. Van Kampen NG (1981) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
11. Grima R (2010) An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *J Chem Phys* 133:035101.
12. Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
13. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
14. Peccoud J, Ycart B (1995) Markovian modeling of gene-product synthesis. *Theor Popul Biol* 48:222–234.
15. Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* 81:3116–3136.
16. Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.
17. Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97:168302.
18. Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol* 54:1–26.
19. Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci USA* 105:17256–17261.
20. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.
21. Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.

22. Phillips R, Kondev J, Theriot J (2009) *Physical Biology of the Cell* (Garland Science, New York).
23. Bernstein JA, Lin PH, Cohen SN, Lin-Chao S (2004) Global analysis of Escherichia coli RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci USA* 101:2758–2763.
24. Lu P, Vogel C, Wang R, Yao X, Marcotte EM (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25:117–124.
25. Bowsher CG (2010) Stochastic kinetic models: Dynamic independence, modularity and graphs. *Ann Stat* 38:2242–2281.
26. Bowsher CG (2011) Information processing by biochemical networks: A dynamic approach. *J R Soc Interface* 8:186–200.
27. Bowsher CG (2011) Automated analysis of information processing, kinetic independence and modular architecture in biochemical networks using MIDIA. *Bioinformatics* 27:584–586.

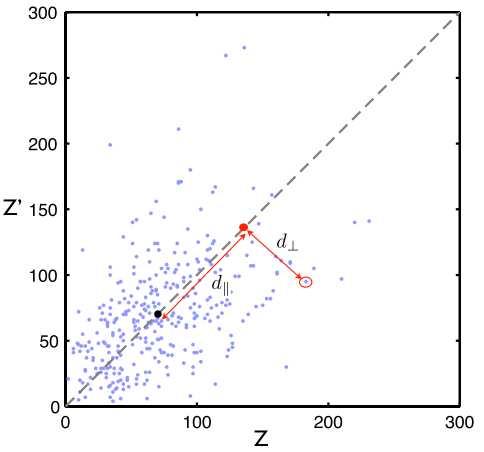

Fig. S1. Typical plot of single-cell measurements of a reporter versus measurements of its conjugate reporter. These data are simulated and are given in numbers of proteins per cell. A typical measurement is highlighted by a red circle, and $d_\perp$ and $d_\parallel$ are shown for this measurement. The diagonal $Z = Z'$ is shown by dashes. The mean $(E[Z], E[Z])$ lies on this diagonal and is shown as a black dot. The point of intersection with the diagonal of the line from (Z, Z') perpendicular to the diagonal is shown as a red dot. This line has a gradient of −1, and the point of intersection is $((Z+Z')/2, (Z+Z')/2)$.

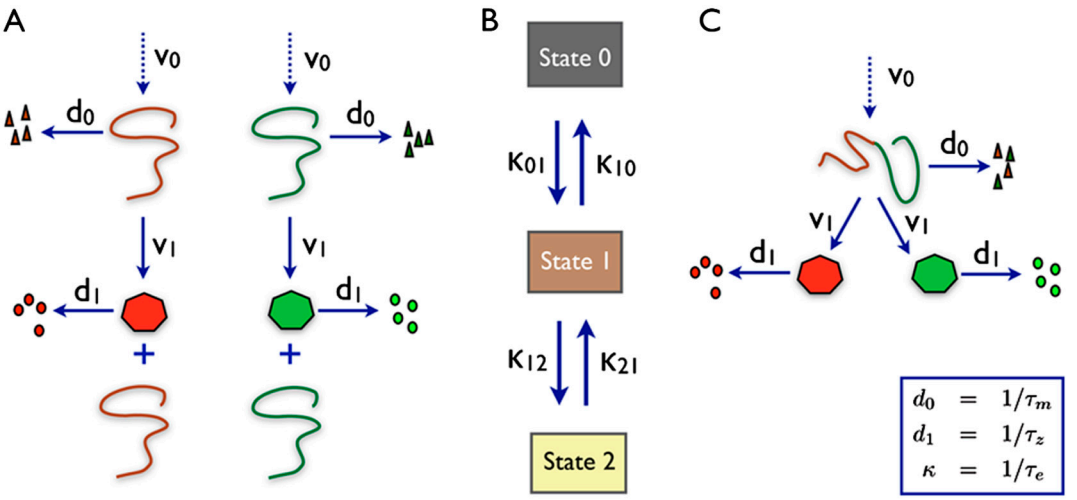

Fig. S2. Reactions for models of gene expression. (*A*) Conjugate reporters given the history of all stochastic processes extrinsic to gene expression. Here, v_0 is the probability of transcription per unit time; v_1 is the probability of translation per unit time per molecule; d_0 is the degradation rate of mRNA per unit time per molecule; and d_1 is the degradation rate of protein per unit time per molecule. (*B*) The local environment is modeled as a Markov chain. It transitions between three states generating extrinsic fluctuations in a parameter that correspondingly transitions between three values. (*C*) A bicistronic reporter for measuring the translational component of variation in gene expression. The inset shows the correspondence between the notation here and that in the main text. We simulated this model (with all κ_{ij} identical and equal to κ) to generate the data for Fig. 2*B*.

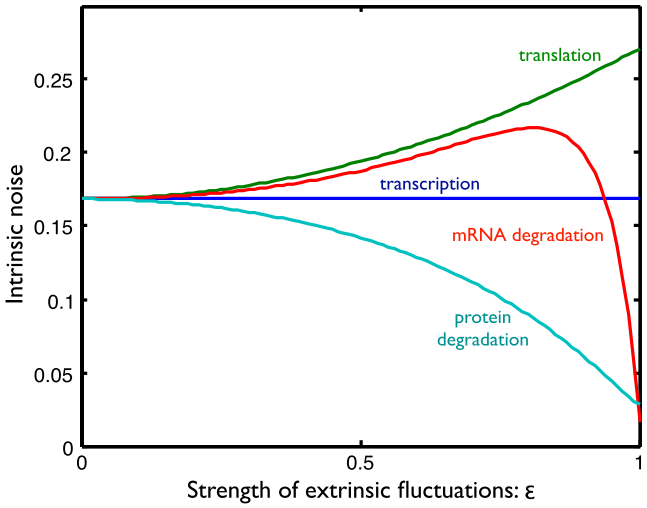

Fig. S3. The dependence of intrinsic noise on extrinsic fluctuations. Exact analytical calculations for intrinsic noise for the model of Fig. S2*A* as the strength of the extrinsic fluctuations in one rate parameter varies. Here, ϵ parametrizes the difference between the extrinsic parameters in each environmental state, and the noise in the extrinsic parameter is $\frac{2\epsilon^2}{3}$. Each curve is marked with the biochemical process that is affected by extrinsic fluctuations. With no extrinsic fluctuations, the intrinsic noise is 0.17 (and equal to the intrinsic noise when only the transcription rate fluctuates). Parameters are the same as those used for the simulations of Fig. 2*B*. For large η_{d_0} (a fluctuating rate of mRNA degradation), the approximation used in Eq. **S61** breaks down, and the intrinsic noise decreases below the value it takes when $\eta_{d_0} = 0$. This nonmonotonic behavior arises because the mean number of proteins increases dramatically as $\epsilon \to 1$.

Table S1. Reaction network and parameter values used in simulations for Fig. 2B

```
variable eta = 0.5
# Z′ reporter
DA + S2 →S2 + DA + MA; v02 = 0.01
DA + S1 → S1 + DA + MA ; v01= v02*(1−eta)
DA + S3 → S3 + DA + MA ; v03 = v02*(1+eta)
MA → MA + A ; v1 = 0.2
MA → null; d0 = 0.0167
A → null; d1 = 0.0017
# Z reporter
# transcription
DB + S2 → S2 + DB + MB; v02
DB + S1 → S1 + DB + MB; v01
DB + S3 → S3 + DB + MB; v03
# translation
MB → MB + B; v1
# degradation
MB → null; d0
B → null; d1
# Z″ (bicistronic) reporter
MB → MB + C; v1
C → null; d1
# state for fluctuations in v0
S1 → S2; k12 = d1/30
S2 → S1; k21 = k12
S2 → S3; k23 = k12
S3 → S2; k32 = k12
INIT
DA = 1 N;
DB = 1 N;
S2 = 1 N;
```