Fundamental trade-offs between information flow in single cells and cellular populations

Ryan Suderman[a,1], John A. Bachman[b,1], Adam Smith[a], Peter K. Sorger[b], and Eric J. Deeds[a,c,d,2]

[a]Center for Computational Biology, University of Kansas, Lawrence, KS 66047; [b]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA 02115; [c]Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047; and [d]Santa Fe Institute, Santa Fe, NM 87501

Edited by Eric D. Siggia, The Rockefeller University, New York, NY, and approved March 2, 2017 (received for review November 4, 2016)

Signal transduction networks allow eukaryotic cells to make decisions based on information about intracellular state and the environment. Biochemical noise significantly diminishes the fidelity of signaling: networks examined to date seem to transmit less than 1 bit of information. It is unclear how networks that control critical cell-fate decisions (e.g., cell division and apoptosis) can function with such low levels of information transfer. Here, we use theory, experiments, and numerical analysis to demonstrate an inherent trade-off between the information transferred in individual cells and the information available to control population-level responses. Noise in receptor-mediated apoptosis reduces information transfer to approximately 1 bit at the single-cell level but allows 3–4 bits of information to be transmitted at the population level. For processes such as eukaryotic chemotaxis, in which single cells are the functional unit, we find high levels of information transmission at a single-cell level. Thus, low levels of information transfer are unlikely to represent a physical limit. Instead, we propose that signaling networks exploit noise at the single-cell level to increase population-level information transfer, allowing extracellular ligands, whose levels are also subject to noise, to incrementally regulate phenotypic changes. This is particularly critical for discrete changes in fate (e.g., life vs. death) for which the key variable is the fraction of cells engaged. Our findings provide a framework for rationalizing the high levels of noise in metazoan signaling networks and have implications for the development of drugs that target these networks in the treatment of cancer and other diseases.

cellular heterogeneity | signal transduction | information theory | apoptosis

Signaling networks allow cells to sense intra- and extracellular concentrations of cytokines, nutrients, ions, and so on and control both discrete and continuous changes in cell state (1–5). Apoptosis and the commitment to cell division are typical of binary responses, whereas directed cell movement and induced gene expression are typical of continuously variable responses (2, 4, 6). Dysregulation of intracellular signaling has been implicated in a wide range of diseases including cancer, chronic inflammation, neurodegeneration, and others (7). Understanding how information is processed in cells will be essential to developing rational therapies for treating such diseases (5, 8). Although signaling networks have long been the subject of intense experimental and theoretical study, it has only recently become possible to monitor signal transduction at the level of individual cells (1, 9, 10). These studies have shown that signaling networks are subject to significant noise, which manifests itself as stochastic fluctuations in the activities of signaling proteins and as cell-to-cell variability among genetically identical cells in a population (1, 3, 8, 11–15). Such noise is ubiquitous and, although its physiological impact remains unclear (16–18), variability has been shown to reduce the effectiveness of drugs that inhibit oncogenic and inflammatory signaling (8).

Information theory (19) provides a powerful analytical framework for quantifying the impact of noise on the ability of a system to transmit information. Levchenko and coworkers pioneered the application of information theory to signaling in mammalian cells (9) with the concentration of an extracellular ligand (e.g., the inflammatory cytokine TNF-α) serving as the input to a (potentially noisy) intracellular signaling network (or channel), leading to a downstream response that can be experimentally measured (the nuclear translocation of NF-κB). They estimated a channel capacity (which is the maximum information a system can transmit) of less than 1 bit for TNF-induced NF-κB translocation to the nucleus based on single-cell data (Table 1, entries 1–4 and 12) (9, 20–22). Because the number of distinct signal values that can be resolved approaches 2^C asymptotically (23), this analysis implies that intracellular signaling networks are barely able to distinguish between the presence or absence of TNF.

More recent research has focused on characterizing strategies that cells might use to achieve higher levels of information transfer. Lee et al. (24) argued that detecting the fold change between a steady-state and induced signal reduces the impact of noise, although we found that fold-change detection in the TNF system still transmits less than 1 bit of information (entry 5, Table 1). Wollman and coworkers (10) recently demonstrated that using multiple time points from the trajectory of a molecular response (e.g., Erk activation over time) significantly increases estimated channel capacities. However, for a cell to use dynamics to increase C requires biochemical circuits for storing and retrieving information, which would themselves be subject to noise (10).

It is difficult to interpret the physiological significance of low channel capacities in published work on signaling because the outputs being measured (e.g., nuclear localization of the NF-κB transcription factor or Erk activation) do not correspond directly to well-defined changes in cell fate (9, 24). We therefore focused on an unambiguous phenotype: life or death as regulated by TNF-related apoptosis-inducing ligand (TRAIL). TRAIL induces apoptosis by binding to cell surface receptors and initiating the formation of death-inducing signaling complexes (DISCs). These

Significance

The function of signal transduction networks is to detect changes in the extracellular environment and combine this with information on intracellular state and thereby control cell-fate decisions. Recent evidence suggests that high levels of biochemical noise in eukaryotic signaling networks interfere with information transmission, making it unclear how cell fate is correctly controlled. Here, we show that high noise levels are advantageous when a system needs to regulate the behavior of populations of cells using noisy biological signals. In contrast, when the key biological unit is a single cell, we show that the impact of noise on cellular responses is much less pronounced. Understanding how noise is generated and exploited advances our understanding of information processing in cells.

Author contributions: R.S., J.A.B., P.K.S., and E.J.D. designed research; R.S., J.A.B., and A.S. performed research; A.S. contributed new reagents/analytic tools; R.S., J.A.B., A.S., P.K.S., and E.J.D. analyzed data; and R.S., J.A.B., P.K.S., and E.J.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]R.S. and J.A.B. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: deeds@ku.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1615660114/-/DCSupplemental.

complexes then initiate a sequence of biochemical events resulting in mitochondrial outer membrane permeabilization (MOMP), activation of the effector caspases (ECs), and cell death (Fig. 1*A*). Dramatic cell-to-cell variability has been observed in the responses of clonal cells to TRAIL (and other death ligands): a subset of cells dies within 2–8 h of ligand exposure, whereas others survive indefinitely. When these survivors are reassayed for TRAIL sensitivity following outgrowth the same fractional killing is observed, showing that variability is a stable property of the cell population. Molecular studies have shown that this variability arises from extrinsic noise in receptor-to-caspase signaling networks (1, 11, 12).

In this paper we show that noise in cell signaling has different effects at the level of single cells and cell populations. When the key physiological output is the behavior of a single cell (as in chemotaxis or mating), high levels of information transfer are observed, implying that noise is likely suppressed (10, 16–18, 24). However, when the key physiological output is the fraction of cells in a tissue or population committing to apoptosis, increased noise reduces information transfer at the level of single cells while increasing it at the level of cell populations. This trade-off between single-cell and population-level information flow is particularly significant when the input signal (e.g., ligand concentration) itself exhibits even small amounts of noise.

Results

Estimating Channel Capacities. The information carried by a channel is quantified by the mutual information, I:

$$I(X;Y) = \sum_X \sum_Y p(x,y) \log \frac{p(x,y)}{p(x),p(y)}, \quad [1]$$

where X is the random variable representing the signal and Y is the variable representing the response (9, 19). The base of the logarithm determines the units of the mutual information: the conventional base 2 quantifies information in "bits" (25). Because the value of I depends on the input distribution, the mutual information of a signaling channel represents a combination of the properties of the signal and the intrinsic limits of the channel itself. Therefore, using mutual information to evaluate information flow in cell signaling networks necessitates an analysis of the properties of input signal distributions in vivo, which are rarely known.

The maximum possible information that a channel can carry, the channel capacity, C, is

$$C = \sup_{p_X(x)} I(X:Y), \quad [2]$$

where the supremum (the least upper bound) is evaluated over all possible choices of the probability distribution of the input. C is an inherent feature of the channel: the larger the value, the more information that a channel can theoretically transmit (9, 19).

Although Eqs. **1** and **2** seem straightforward, estimation of mutual information and channel capacity from experimental data is a nontrivial challenge. Recent algorithms make it possible to estimate the channel capacity between cellular signals and the downstream responses they control (9). These approaches use empirical dose–response data to estimate $p(y|x)$ and then search over a finite set of input distributions $P = \{p_X(x)\}$ to find the one that maximizes I. Although the resulting values are reported as "channel capacities" (9, 10), with a finite set of possible input signals S and a finite set of probability distributions P, one cannot find the supremum as defined in Eq. **2** (25, 26). We therefore refer to the maximum mutual information calculated in this manner as the "estimated channel capacity given S and P" ($\hat{C}_{S,P}$) or simply the "estimated channel capacity" ($\hat{C}$).

Table 1. Estimated channel capacity for experimental data

Signal, response	Signal	Response	$\hat{C}_{S,P}$ (bits)	Data source	Calculation source
Molecular, molecular	1. TNF	NF-κB	0.92 ± 0.01	9	9
	1.1. TNF	ATF-2	0.85 ± 0.02	9	9
	1.2. TNF	NF-κB & ATF-2	1.05 ± 0.02	9	9
	2. PDGF	NF-κB	0.67 ± 0.01	9	9
	2.1. PDGF	ATF-2	0.74 ± 0.01	9	9
	2.2. PDGF	NF-κB & ATF-2	0.81 ± 0.02	9	9
	3. EGF	Erk (fold-change)	0.60 ± 0.03	21	9
	4. UDP	Peak Ca^{2+}	1.22 ± 0.03	20	9
	4.1. UDP	Integrated Ca^{2+}	1.07 ± 0.02	20	9
	5. TNF	A20 transcripts	0.84 ± 0.10	24	This work
	6. TRAIL	Casp-8 activity	1.01 ± 0.01	This work	This work
	7. TRAIL	Casp-8 activity (live cells)	1.01 ± 0.01	This work	This work
	8. TRAIL	Casp-3 activity	0.56 ± 0.01	This work	This work
	9. Casp-8 activity	Casp-3 activity	1.23 ± 0.01*	This work	This work
	10. Casp-8 activity	Cell decision	0.63 ± 0.01	This work	This work
	11. α-Factor	*pFUS1*-GFP	2.26 ± 0.04	6	This work
Position, molecular	12. Embryo perimeter	Phosphorylated Erk	1.61 ± 0.05	22	9
Position, motion	13. Bacterium	Neutrophil motion	1.82 ± 0.11		This work
	14. cAMP	*Dictyostelium* motion	2.19 ± 0.08	Firtel laboratory	This work
Molecular, population	15. TRAIL	% dead (HeLa; resampled)	2.44 ± 0.02	This work	This work
	16. TRAIL	% dead (HeLa; FACS)	3.41 ± 0.03	This work	This work
	17. TRAIL	% dead (MCF10A)	3.38 ± 0.01	This work	This work

The estimated channel capacity for population-level response in HeLa cells was calculated using 1,000 cells per TRAIL concentration and all population-level channel capacities were calculated using 100 independent populations. Ranges on values in the table represent 95% confidence intervals, calculated using the robust variance estimator (see *Materials and Methods* and *SI Appendix*). Bootstrap estimates of the confidence intervals result in similar values (Table S1 and *SI Appendix*).

*Indicates that the value is an underestimate (the optimal number of bins for estimation was not reached due to computational limitations).

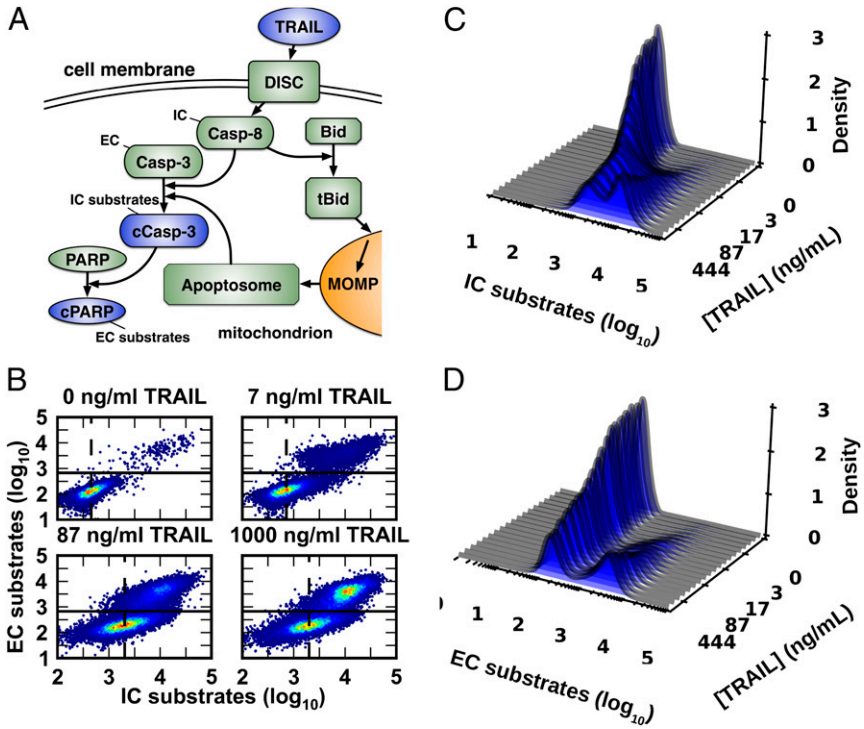

Fig. 1. Cell-to-cell variability in response to a range of TRAIL doses. (*A*) TRAIL activates the IC Casp-8 via DISCs. Active Casp-8 then activates the EC Casp-3 via two mechanisms: direct cleavage and MOMP, which induces formation of the apoptosome, another activator of Casp-3 (13). (*B*) Measurement of cleaved EC and IC substrates by flow cytometry (13) shows that HeLa cells have a highly variable response to TRAIL across a wide range of doses (n = 60,000 cells per TRAIL dose). The solid line is the minimum density in the bimodal EC response (~2.8 in $\log_{10}$ units) and acts as a threshold for apoptosis, whereas the dashed line marks the average IC response for nonapoptotic cells. We used kernel density estimators to estimate TRAIL-dependent response distributions for IC (*C*) and EC (*D*) activity.

To facilitate comparison between our calculations and those performed by Cheong et al. (9) we designed a software package to estimate mutual information based on the binning procedure they applied in their work (see *Materials and Methods* and *SI Appendix* for further details). This software is freely available (https://github.com/ryants/EstCC).

Individual Cells Responding to TRAIL Exhibit a Low Channel Capacity. To estimate the channel capacity of the extrinsic apoptosis signaling cascade, HeLa cells were treated with TRAIL for 11 h over a range of ligand concentrations from sub- to superphysiological, and molecular responses in single cells were then measured by flow cytometry (12). The level of cleaved caspase-3 (cC3) served as a measure of the time-integrated activity of initiator caspases (ICs) and cleaved PARP (cPARP) served as a measure of downstream EC activity (Fig. 1*A*). Previous studies have shown that TRAIL exposure results in a dose-dependent increase in IC activity that varies significantly from cell to cell; in any single cell, when IC activity exceeds a threshold set by antiapoptotic Bcl-2 proteins, ECs are activated and the cell proceeds inexorably to death (Figs. 1 *B–D*) (1, 12).

Using the distributions of IC and EC activity in single cells, we calculated an estimated channel capacity between TRAIL dose and IC activity of $\hat{C}$ ~1.01 bits and between TRAIL and EC activity of $\hat{C}$ ~0.56 bits (entries 6 and 8, Table 1). Previous studies in our groups using identical experimental methods show high correlations between technical and biological repeats, suggesting that low estimated channel capacities are unlikely to reflect noise in the instrument or errors in experimental technique (27). We considered the possibility that dead or dying cells (those with EC levels above the death threshold) would exhibit apparently increased levels of IC activity due to feedback by ECs (28), masking or degrading the signal contributed directly by the upstream TRAIL/receptor axis. We therefore estimated channel capacity between TRAIL and IC activity in surviving cells that did not activate ECs and observed similarly low values (entry 7, Table 1 and *SI Appendix*, Figs. S3 and S4).

To determine whether noise was introduced primarily by the upstream (receptor to IC) or downstream (IC to EC) step, we calculated the channel capacity between IC and EC activity and obtained $\hat{C}$ = 1.23 bits (entry 9, Table 1). The larger $\hat{C}$ for the downstream step compared with both the upstream step (TRAIL to IC, $\hat{C}$ = 1.01 bits) and the overall process (TRAIL to EC, $\hat{C}$ = 0.56 bits) indicates that, whereas activation of ICs by a given TRAIL dose is highly variable among cells, IC activity itself is predictive of cellular responses, consistent with our previous work (1, 29).

Populations of Cells Responding to TRAIL Exhibit High Channel Capacity. When we examined the channel capacity between TRAIL dose and phenotypic response at the population level we obtained a very different result. The combination of noise and a threshold causes the fraction of cells that die to vary smoothly with TRAIL concentration (1, 13–15, 30). For either of two cell types (transformed HeLa and nontransformed MCF10a cells) we found that the fraction of cells surviving exposure to TRAIL gradually decreased as the concentration of ligand increased over a 10^3-fold range, with comparatively little variance between replicate experiments (Fig. 2). As a result, the estimated channel capacity between TRAIL dose and the fraction of cells undergoing apoptosis was much higher than that observed for the molecular response in single cells, $\hat{C}$ ~3.4–4 bits depending on the population size (entries 15 and 16 in Table 1 and Fig. 3). This finding suggests that low information transfer at the single-cell level is accompanied by high rates of transfer in cell populations, allowing TRAIL dose to control the fraction of cells that die without precisely specifying which cells in the population cross the IC activity threshold.

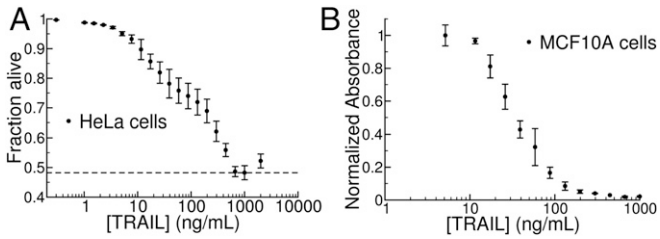

Fig. 2. Population-level dose–response relationship for TRAIL-mediated apoptosis. (*A*) We used the threshold described in Fig. 1*B* to map data from HeLa cells to fractional survival at varying TRAIL doses. We recorded a maximal effect at [TRAIL] = 1,000 ng/mL (indicated by the dashed line); higher doses of TRAIL lead to less fractional killing in a "ligand squelching" effect that we have consistently observed for this system. Because the channel capacity represents a supremum over all possible probability distributions of input signals, we removed the final point ([TRAIL] = 2,000 ng/mL) from our analysis without loss of generality. Error bars indicate sample standard deviation across three replicates of 20,000 cells each. (*B*) Fraction of MCF10A cells surviving TRAIL treatment as assayed by methylene blue staining (*SI Appendix*) (12) shows a graded response similar to that of the HeLa population in *A*.

A Trade-Off Between Single-Cell and Population-Level Information Transfer. To understand how noise in single cells contributes to high channel capacity at a population level (3, 13–15) we created a series of simple models of intracellular signaling. In the first model, the response R of individual cells to a signal dose S is related by a Hill function modified to account for noise:

$$R = (R_{\max} - R_{\min}) \cdot \frac{S^n}{S^n + K^n} + R_{\min} + \varepsilon, \quad [3]$$

where K is the concentration of an input ligand that results in a half-maximal response, n is the Hill exponent (a measure of ultrasensitivity in the dose–response relationship), $R_{\min}$ and R_{max} represent the range of average responses, and ε is a noise term sampled from a Gaussian distribution with mean $\mu = 0$ and variable standard deviation, σ (12). We used this model to generate a series of single-cell dose–response curves with different levels of noise (Fig. 3*A*) and estimated $\hat{C}(S;R)$ using the same approach we applied to the experimental data (Figs. 1 and 2). To map individual cell responses to a binary decision, we introduced a threshold θ (dashed line in Fig. 3*A*): individual cells with $R < \theta$ survive, whereas those with $R \geq \theta$ die. This mimics the MOMP threshold set by antiapoptotic Bcl-2 proteins (12). Using this threshold, we can obtain population-level transitions in the percentage of cells that die as a function of S (Fig. 3*B*).

For simulated populations of $n = 10^2$ to 10^4 cells, this model revealed a striking trade-off between the estimated channel capacity for single cells and populations. In this case, we used an evenly spaced set of signal values characterized by a minimal signal resolution ΔS, similar to experimental systems (Figs. 1 and 2) (9, 10). When noise is low, the entire population-level transition happens between two neighboring S values, leading to a step-like response in the population (Fig. 3 *A* and *B*, blue). At higher levels of noise, the response of individual cells becomes more variable, and the fraction of cells that die gradually increases (Fig. 3 *A* and *B*, black and red) (1, 13–15, 30). Low noise thus leads to high estimated channel capacities between S and R measured at the single-cell level (5 bits or more) but low estimated channel capacities between S and the fraction of cells that die (~1 bit or less, Fig. 3*C*).

Although the trade-off observed in Fig. 3*C* is suggestive, it is important to note that the minimal step size ΔS used for this calculation is arbitrary. If, instead, we use the principle of gain adaptation (i.e., adjusting ΔS so that signal values are finely sampled when R varies most rapidly) (25, 26), we find that population-level channel capacities become insensitive to noise below a certain ΔS value (*SI Appendix*, Figs. S12 and S13). To better characterize the relationship between single-cell noise and population-level information transfer we developed an analytical framework for the S/R model (*SI Appendix*, sections 5–7). Although the channel capacities for the original model with Gaussian-distributed noise could not be expressed in closed form, a simplified model with a uniform distribution of noise proved to be analytically tractable. In this simplified model, if we consider the limit of continuous signals (i.e., $\Delta S \to 0$), we find that the estimated population-level channel capacity follows:

$$C \sim \frac{1}{2} \log N \quad [4]$$

for any nonzero value of noise below a certain threshold (*SI Appendix*, section 7).

The question then becomes: why might cells evolve networks with high levels of noise to maximize population-level information flow, if even a very small amount of noise would suffice to accurately control the population? The answer to this question likely lies in the fact that, at the local level at which they act in paracrine regulation, cytokines such as TRAIL are not present at uniform, precise concentrations. Rather, they are synthesized and catabolized locally, and therefore subject to Poisson-style birth–death noise (31, 32). To get a sense of the scale of these fluctuations, consider a tissue compartment with volume similar to that of an intestinal crypt (33, 34). In this case, the concentrations of TRAIL used in our in vitro experiments correspond to $\sim 10^2$ to 10^4 molecules per compartment. Cells must maintain these concentrations by synthesizing TRAIL to make up for cytokine loss from the compartment due to degradation or dilution, and so we can expect relative fluctuations in TRAIL concentrations to approach 1–10% (*SI Appendix*).

We found that, even with 1% fluctuation in the input signal, population-level estimated channel capacities drop dramatically at low levels of noise, even when ΔS is taken to be arbitrarily small (Fig. 3*D*). This occurs because when single-cell noise levels are low the population-level transition becomes very sharp (Fig. 3*B*), and ΔS must be very small to capture that transition (*SI Appendix*). In this case, noise in the signal generates "confusion" from nearby signals, which makes it difficult to control the population. In contrast, if levels of noise in individual cells are relatively high, the population-level transition occurs across a wide range of signal values, and the impact of small fluctuations in signal value becomes less important and population level control increases.

As mentioned above, our analytical calculations indicated that population-level channel capacity should increase as 1/2 log N. To examine the effect of population size on estimated channel capacity using experimental data, we randomly sampled subpopulations of the $\sim 6 \times 10^4$ HeLa cells analyzed at each TRAIL dose. We found exactly the predicted dependence of population-level information transfer on population size in both our experimental data and our Gaussian model (Fig. 3*E*).

Low Channel Capacities Likely Do Not Represent Intrinsic Biophysical Limits. We next explored whether the noise observed in TNF and TRAIL signaling represents a physical limit for eukaryotic signal transduction circuits. To do this, we considered two cases in which individual cells (rather than cell populations) must respond accurately to environmental stimuli. During *Saccharomyces cerevisiae* mating, haploid **a** and **α** cells must determine whether a suitable mating partner is sufficiently close for conjugation to be successful. Cells sense the local concentration of the mating pheromone α factor via a G protein-coupled cell surface receptor and a downstream MAP kinase cascade; when a suitable partner is available for conjugation they reorient their cytoskeleton and initiate a complex transcriptional program (4, 6). The decision to mate results in cell-cycle arrest (35) and is highly consequential for yeast cell fate, and so we expect there would evolutionary pressure for individual cells

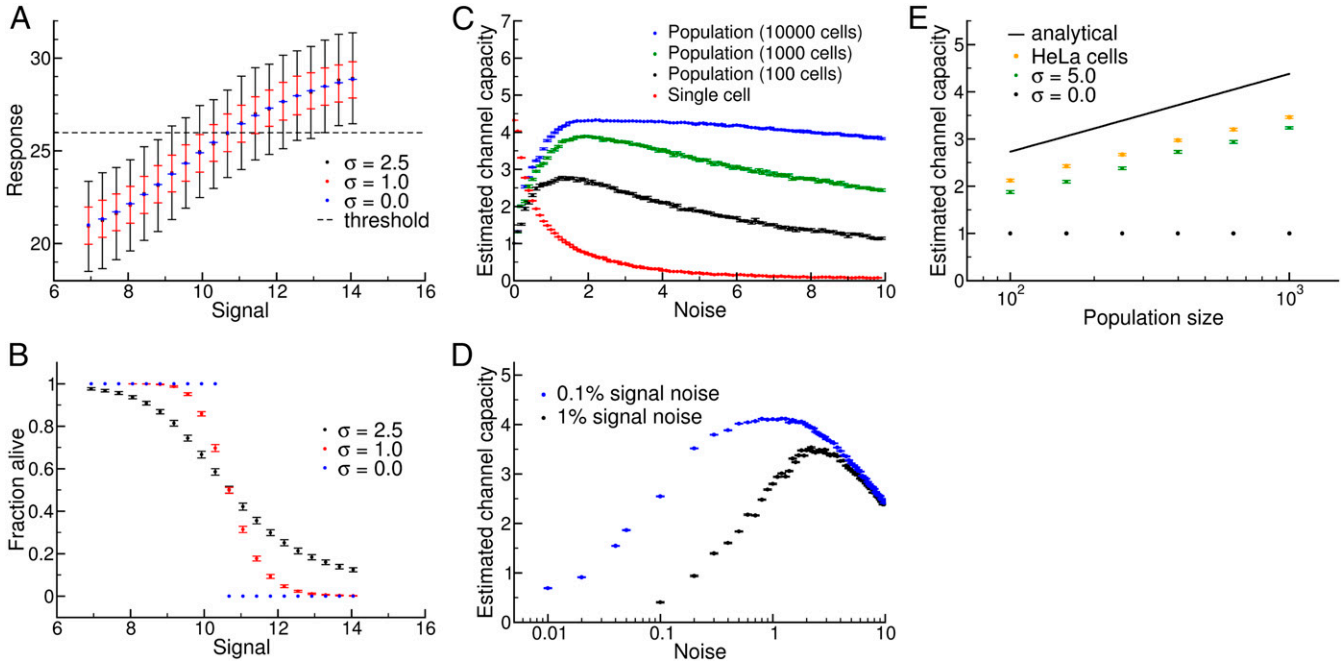

Fig. 3. Relationship between single-cell and population-level channel capacity. (*A*) Single-cell dose–response behavior for the model described in Eq. **3** (see *SI Appendix* for model parameters). The mean response and standard deviation of 1,000 independent simulated "cells" is shown for various noise values, relative to a cell death threshold (dashed line). (*B*) Population dose–response behavior from $P = 100$ independent populations with $n = 1{,}000$ cells per signal. Individual cells' response map to either death or survival according to the threshold in *A*. (*C*) Increasing noise decreases information transmission in single cells and simultaneously increases the population-level channel capacity up to an optimal noise value. (*D*) Estimated channel capacity for the model in Eq. **3**, but extended to include noise in the signal. Here, the *x* axis is the noise in the response as in *C* and the points represent two levels of relative signal noise (1% and 0.1%; *SI Appendix*). Even with arbitrarily high signal resolution, relatively high levels of single-cell noise are required to compensate for realistic levels of noise in signal values. (*E*) Scaling of channel capacity with population size. The black curve is our analytical prediction, $C \sim 1/2 \log N$. The orange points were obtained by resampling our experimental data for HeLa cells at the indicated population sizes, and the green points are results from our Gaussian signal-response model. Note that $\hat{C}$ is slightly lower in both cases, and there is a slight deviation in the scaling relationship at higher values of *N*; both of these phenomena derive from limited sampling of signals (i.e., a particular ΔS; *SI Appendix*).

to acquire accurate information about the availability of a mating partner. We calculated $\hat{C}$ ~2.26 bits between pheromone dose and transcriptional output as measured by a fluorescent reporter (entry 11, Table 1). This value is achieved even though we consider only single-cell data on the status of the pheromone-sensing network (6), and not the prior history of signaling, as did Doncic et al. (16–17) and Doncic and Skotheim (18); consideration of this historical information would likely increase our channel capacity estimate (10).

Another situation in which individual cells are the key biological actors is eukaryotic chemotaxis. We therefore analyzed a classic movie of a human neutrophil "hunting" a bacterial cell and a movie of a single *Dictyostelium* cell responding to cAMP emanating from a micropipette (https://www.youtube.com/watch?v=Kb-m1uDoWfU and Movie S1, respectively) (36). Because migrating cells are polar, it is possible to define a cell-based coordinate system using standard tracking software (CellTrack) (37). Following the lead of others working on distributions of directional movement (38), we defined the input as the angle between the chemoattractant (bacterium or micropipette) and the cell axis and the output as the angle of the subsequent motion (*SI Appendix*). For neutrophils and *Dictyostelium* we computed $\hat{C}$ ~1.8 and C ~2.2 bits (entries 13 and 14, Table 1), which is almost certainly a lower bound given that we simplify a 3D problem as a 2D search (*Discussion*). Nonetheless, we conclude that signaling networks in single cells can likely encode more than 2 bits of information.

Discussion

Our findings touch on two distinct and complementary aspects of information transfer in signal transduction: single-cell and population-level information processing. In the case of regulatory networks that control apoptosis, the key physiological variable is the fraction of cells responding at a given dose (14, 30). In this case, low estimated channel capacity at a single-cell level ($\hat{C} < 1$) results in high capacity at a population level ($\hat{C}$ ~3–4). We might therefore expect that when a binary cell-fate decision is the output of a signaling system, heterogeneity in the population induced by noise allows precise control of the fraction of cells above (or below) a decision threshold (1, 2, 13–15, 30).

Our analysis indicates that variability in the input signal is a key factor in understanding the interplay between single-cell noise and population-level control. With cell culture experiments in which a ligand is added at specified concentration, it is reasonable to assume that the input signal is uniform among cells. In an organism, however, the distribution of the signal is likely to be much more complex. The levels of soluble ligands are subject to fluctuation in rates of synthesis and degradation in tissue compartments of limited size. From the point of view of calculating channel capacities, a key insight of our work is that noise in the input signal limits the ability to achieve high levels of population-level control when there is high fidelity in single-cell signaling. In contrast, increasing single-cell noise allows cells to optimize population control in the face of an unpredictable and noisy environment. Although our results are suggestive, further experimental work will be needed to characterize the signal distributions that control population-level responses in vivo.

A different situation arises when individual cells must precisely resolve signals to make decisions in a continuous response space. For eukaryotic chemotaxis, we found that the single-cell channel capacity is significantly higher than has hitherto been observed ($\hat{C}$ ~2, Table 1, entries 13 and 14). As mentioned above, we expect that this value is likely a lower bound; for instance, the

data for the *Dictyostelium* calculation ($\hat{C}$ ~2.2, Table 1, entry 14) exhibits tight distributions around ~6 input angles, producing a maximum input entropy (and thus an upper limit on $\hat{C}$) of slightly less than $\log_2$ (6) = 2.6 bits. Because the estimated channel capacity is close to this value, we suspect that the calculated spatial channel capacity in this case is limited by insufficiently fine sampling of signaling space (i.e., the relative angle between the source of the signal and the cell).

Another prototypical example of a cell-fate decision that is critical for an individual cell but not a population is cell-cycle arrest by yeast in response to pheromone secreted by cells of the opposite mating type (35). We found that the pheromone signaling network can transmit over 2 bits of information to transcriptional responses in single yeast cells (Table 1, entry 11). Interestingly, Doncic et al. (17) thoroughly characterized the molecular mechanisms controlling this decision-making process, and they found that events such as nuclear export of Whi5 accurately predict whether a cell will arrest or commit to division. The decision to arrest the yeast cell cycle thus involves signaling motifs that integrate information over multiple time scales, from rapid sensing of pheromone gradients on the order of seconds to signaling events from multiple past generations (a timescale of many hours), enhancing control over cell-fate decisions (16, 18).

These findings suggest that low channel capacities at a single-cell level ($\hat{C} < 1$) do not reflect an inherent limit in the biochemistry of signal transduction, but rather a natural trade-off between the knowledge that individual cells have about their environment and the ability of multicellular organisms to control responses reliably at the population level. With respect to noise levels in these systems, two nonexclusive possibilities exist. The first is that networks that control cellular populations simply exploit noise that arises from stochastic fluctuations in transcription, protein synthesis, and related processes whereas chemotactic and mating networks have evolved to suppress it (10, 16–18). The second is that some signaling networks have actually evolved higher levels of noise than the underlying biophysics might dictate (39–42). Expression of T-cell receptors seems to be one case in which regulatory networks are structured to increase cell-to-cell variability (13). In either case, the physiological importance of noise may explain why drugs that target cellular decision networks have difficulty eliciting complete population-level responses (8). Understanding and ultimately exploiting biological noise is thus likely to be as important for therapy as it is for understanding metazoan signaling.

Materials and Methods

Experimental Methods. See *SI Appendix*, section 8 for details on the experimental methods used in this work.

Calculating Mutual Information. Our method for calculating channel capacities was based on the approach described in Cheong et al. (9). As in their case, we randomly resampled the data across a range from 60 to 95% to extrapolate the mutual information at infinite sample size. The error ranges reported in Table 1 are the 95% confidence intervals for the estimate of mutual information extrapolated to infinite sample size. The source code can be found at https://github.com/ryants/EstCC and a complete description of the estimation procedure is given in *SI Appendix*.

ACKNOWLEDGMENTS. The authors would like to thank Tom Kolokotrones and Walter Fontana for many helpful discussions. This work was funded by NIH Grant P50-GM107618, Army Research Office Grant W911NF-14-1-0397 (to P.K.S.), and National Science Foundation Grant MCB-1412262 (to E.J.D.).

1. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459:428–432.
2. Chen J-Y, Lin J-R, Cimprich KA, Meyer T (2012) A two-dimensional ERK-AKT signaling code for an NGF-triggered cell-fate decision. *Mol Cell* 45:196–209.
3. Balázsi G, van Oudenaarden A, Collins JJ (2011) Cellular decision making and biological noise: From microbes to mammals. *Cell* 144:910–925.
4. Suderman R, Deeds EJ (2013) Machines vs. ensembles: Effective MAPK signaling through heterogeneous sets of protein complexes. *PLOS Comput Biol* 9:e1003278.
5. Rowland MA, Fontana W, Deeds EJ (2012) Crosstalk and competition in signaling networks. *Biophys J* 103:2389–2398.
6. Bashor CJ, Helman NC, Yan S, Lim WA (2008) Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* 319:1539–1543.
7. Sebolt-Leopold JS, Herrera R (2004) Targeting the mitogen-activated protein kinase cascade to treat cancer. *Nat Rev Cancer* 4:937–947.
8. Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK (2013) Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat Chem Biol* 9:708–714.
9. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. *Science* 334:354–358.
10. Selimkhanov J, et al. (2014) Systems biology. Accurate information transmission through dynamic biochemical signaling networks. *Science* 346:1370–1373.
11. Flusberg DA, Roux J, Spencer SL, Sorger PK (2013) Cells surviving fractional killing by TRAIL exhibit transient but sustainable resistance and inflammatory phenotypes. *Mol Biol Cell* 24:2186–2200.
12. Albeck JG, et al. (2008) Quantitative analysis of pathways controlling extrinsic apoptosis in single cells. *Mol Cell* 30:11–25.
13. Feinerman O, Veiga J, Dorfman JR, Germain RN, Altan-Bonnet G (2008) Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science* 321:1081–1084.
14. Ahrends R, et al. (2014) Controlling low rates of cell differentiation through noise and ultrahigh feedback. *Science* 344:1384–1389.
15. Shalek AK, et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510:363–369.
16. Doncic A, et al. (2015) Compartmentalization of a bistable switch enables memory to cross a feedback-driven transition. *Cell* 160:1182–1195.
17. Doncic A, Falleur-Fettig M, Skotheim JM (2011) Distinct interactions select and maintain a specific cell fate. *Mol Cell* 43:528–539.
18. Doncic A, Skotheim JM (2013) Feedforward regulation ensures stability and rapid reversibility of a cellular state. *Mol Cell* 50:856–868.
19. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423.
20. Bao XR, Fraser IDC, Wall EA, Quake SR, Simon MI (2010) Variability in G-protein-coupled signaling studied with microfluidic devices. *Biophys J* 99:2414–2422.
21. Cohen-Saidon C, Cohen AA, Sigal A, Liron Y, Alon U (2009) Dynamics and variability of ERK2 response to EGF in individual living cells. *Mol Cell* 36:885–893.
22. Coppey M, Boettiger AN, Berezhkovskii AM, Shvartsman SY (2008) Nuclear trapping shapes the terminal gradient in the Drosophila embryo. *Curr Biol* 18:915–919.
23. Cover TM, Thomas JA (1991) *Elements of Information Theory* (Wiley, New York).
24. Lee REC, Walker SR, Savery K, Frank DA, Gaudet S (2014) Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Mol Cell* 53:867–879.
25. Levchenko A, Nemenman I (2014) Cellular noise and information transmission. *Curr Opin Biotechnol* 28:156–164.
26. Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. *Neuron* 26:695–702.
27. Gaudet S, Spencer SL, Chen WW, Sorger PK (2012) Exploring the contextual sensitivity of factors that determine cell-to-cell variability in receptor-mediated apoptosis. *PLOS Comput Biol* 8:e1002482.
28. Wieder T, et al. (2001) Activation of caspase-8 in drug-induced apoptosis of B-lymphoid cells is independent of CD95/Fas receptor-ligand interaction and occurs downstream of caspase-3. *Blood* 97:1378–1387.
29. Roux J, et al. (2015) Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold. *Mol Syst Biol* 11:803.
30. Miller M, et al. (2012) Modular design of artificial tissue homeostasis: Robust control through synthetic cellular heterogeneity. *PLOS Comput Biol* 8:e1002579.
31. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
32. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–418.
33. Lencer WI, et al. (1997) Induction of epithelial chloride secretion by channel-forming cryptdins 2 and 3. *Proc Natl Acad Sci USA* 94:8585–8589.
34. van der Wath RC, Gardiner BS, Burgess AW, Smith DW (2013) Cell organisation in the colonic crypt: A theoretical comparison of the pedigree and niche concepts. *PLoS One* 8:e73204.
35. Hartwell LH, Culotti J, Pringle JR, Reid BJ (1974) Genetic control of the cell division cycle in yeast. *Science* 183:46–51.
36. Janetopoulos C, Firtel RA (2008) Directional sensing during chemotaxis. *FEBS Lett* 582:2075–2085.
37. Sacan A, Ferhatosmanoglu H, Coskun H (2008) CellTrack: An open-source software for cell tracking and motility analysis. *Bioinformatics* 24:1647–1649.
38. Burov S, et al. (2013) Distribution of directional change as a signature of complex dynamics. *Proc Natl Acad Sci USA* 110:19689–19694.
39. Shahrezaei V, Swain PS (2008) The stochastic nature of biochemical networks. *Curr Opin Biotechnol* 19:369–374.
40. Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Phys Rev Lett* 97:168302.
41. Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–362.
42. Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* 81:3116–3136.

Supporting Information

Suderman et al. 10.1073/pnas.1615660114

Movie S1. *Dictyostelium* chemotaxis in response to the chemoattractant cAMP. The *Dictyostelium* cell responds to changes in spatial information (the location of the micropipette) by extending pseudopodia and migrating in the direction of the chemoattractant source. Movie provided courtesy of Susan Lee and Richard Firtel, Section of Cell and Developmental Biology, Division of Biological Sciences, University of California, San Diego, La Jolla, CA.

Movie S1

Other Supporting Information Files

SI Appendix (PDF)

Fundamental Trade-offs between Information Flow in Single Cells and Cellular Populations

Supporting Information

Ryan Suderman[1,*], John A. Bachman[2,*], Adam Smith[1], Peter K. Sorger[2] and Eric J. Deeds[1,3]

[1]Center for Computational Biology, The University of Kansas, 2030 Becker Dr., Lawrence, KS 66047
[2]Department of Systems Biology, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115
[3]Department of Molecular Biosciences, The University of Kansas, 1200 Sunnyside Ave., Lawrence, KS 66047
[*]These authors contributed equally to this work

Email: Eric J. Deeds - deeds@ku.edu;

Contents

1 Information Theory Calculations

1.1 Mutual Information

The mutual information between two random variables representing a signal, S, and a response, R, is defined as:

$$I(S;R) = \int_S \int_R p(s,r) \log \frac{p(s,r)}{p(s)p(r)} \mathrm{d}s\mathrm{d}r,$$

where S is a random variable representing the input signal, R a random variable representing the response, $p(s,r)$ is the joint probability distribution for some combination of s and r values, and $p(s)$ and $p(r)$ are the corresponding marginal distributions (1). One of the major difficulties in calculating this quantity from experimental data is the fact that the continuous probability density functions defined above must be estimated on the basis of an inherently discrete data set. As a result, a number of approaches have been developed to obtain unbiased estimates of the mutual information with varying degrees of accuracy (2).

In order to facilitate comparison with earlier results, we employed the same strategy used by Cheong *et al.* (3). This strategy has two main components. First, one defines a set number of "bins" in both the signal values s and response values r. In cases where one is measuring the molecular response of individual cells to a given signal (e.g. nuclear localization of NF-κB upon treatment with TNF-α, (3)), there are a small number of ligand concentrations used to treat the cells, resulting in a natural discretization of the S variable and a total of S_B bins of signal values. One defines a number of bins for the response (R_B), and uses these bins to estimate the probability of observing some response bin given some signal bin (*i.e.* $p(r|s)$). A linear extrapolation procedure is then used to estimate the mutual information one would obtain if there were an infinite amount of data in the data set. This extrapolation procedure is described in greater detail in section 1.1.2 below.

One issue with this approach, however, is that the number of bins into which the signal and response values should be divided is not well-defined; using a larger number of bins generally increases the estimated amount of information (3). To combat the potential for overestimation of the mutual information, the second phase of the procedure involves varying the total number of bins in the response variable (and, when appropriate, in the signal value as well) and estimating I for both the experimental data and a set of randomized replicates of the data. This allows one to choose a bin size that maximizes I for the real data while still estimating 0 information for the randomized versions. This element of the procedure is detailed in section 1.1.3. Estimates of the mutual information based on this approach can subsequently be used to calculate the channel capacity by finding the input distribution that maximizes I (1) (section 1.2).

1.1.1 Calculating the mutual information

To calculate the mutual information from our finite data sets, we first created a "contingency table" K based on the data: the rows of this matrix represent the various signal bins, and the columns are the various response bins. Each entry in the matrix is the number of observations from the data that correspond to that particular signal-response pair. The contingency table for a particular experiment might look something like this:

$$K = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{ccccc} r_1 & r_2 & r_3 & r_4 & r_5 \end{array} \\ \begin{pmatrix} 6 & 1 & 0 & 0 & 0 \\ 0 & 3 & 4 & 0 & 0 \\ 0 & 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 2 & 5 \end{pmatrix} \end{array}$$

Note that the above table is meant only as an example, and does not contain actual data. One can use the contingency table to calculate the mutual information in terms of the marginal and conditional entropies:

$$I(S;R) = H(S) - H(S|R)$$

$$I(S;R) = -\sum_i^{S_B} p(s_i) \log p(s_i) - \sum_i^{S_B} \sum_j^{R_B} p(s_i, r_j) \log \frac{p(r_j)}{p(s_i, r_j)}$$

where i ranges over the signal bins and j over the response bins in the contingency table (recall that S_B and R_B are the total number of signal and response bins, respectively). Since each entry in the contingency table can be naturally considered a conditional probability, it is helpful to rewrite this equation as:

$$I(S;R) = -\sum_i^{S_B} p(s_i) \log p(s_i) + \sum_j^{R_B} p(r_j) \sum_i^{S_B} p(r_j|s_i) \log p(r_j|s_i).$$

We can then calculate the frequencies from the contingency table entries and substitute these values into the equation. We define N_T as the sum over all entries in the table. Since each entry of the matrix, k_{ij}, is the number of instances of signal i that resulted in response j, we can define the total number of observations corresponding to a given signal bin i as $k_{s,i} \equiv \sum_j^{R_B} k_{ij}$. Similarly, we can define the total number of times any particular response bin j was observed as $k_{r,j} \equiv \sum_i^{S_B} k_{ij}$. Given these definitions, we can calculate the mutual information using the following equation:

$$I(S;R) = -\sum_i^{S_B} \frac{k_{s,i}}{N_T} \log \frac{k_{s,i}}{N_T} + \sum_j^{R_B} \frac{k_{r,j}}{N_T} \sum_i^{S_B} \frac{k_{ij}}{N_T} \log \frac{k_{ij}}{N_T}. \tag{1}$$

Equation 1 is used whenever a particular value of I is calculated in the estimation procedure described below (2; 3).

1.1.2 Removing bias due to finite sample size

Although it is straightforward to use equation 1 to calculate the mutual information, the fact that there are a finite number of data points in the contingency table (N_T) can introduce biases into the calculation. To estimate this bias, one can create a smaller data set with N'_T points ($N'_T < N_T$) and calculate I. As has been observed previously (3), as N'_T decreases, bootstrap replicates of the data generate higher values of I. This results in a roughly linear decrease in I as the inverse sample size decreases (Figure S1).

To correct for this bias, we used the linear extrapolation procedure employed in Cheong *et al.* and

other previous studies (3; 4). We chose a total of 5 sample sizes starting from 60% of the original data with uniform increments in inverse sample space until reaching the size of the original data set. Our procedure then involves calculating the joint frequency distribution for the original data set and then randomly sampling the specified number of values over 20 independent replicates from the joint frequency distribution. Note that we re-sample the data *with replacement*, which allows us to generate these 20 replicates accross the entire range of N_T values. Sampling with replacement also allows us to find the signal distribution that maximizes $I(S;R)$ as described in section 1.2 below. We used these randomly sampled data sets to generate new contingency tables, and using equation 1 we calculated the distribution of I across the 20 replicates. We then performed a linear regression of the I *vs.* $1/N_T$ relationship (*e.g.* the straight lines in Figure S1). The y-intercept of these lines represents the extrapolation to an infinite data set (*i.e.* $N_T \to \infty$ implies $1/N_T \to 0$). All the channel capacities calculated in this work (*e.g.* those reported in Table 1 of the main text) were obtained from these y-intercepts. The errors reported for these values in Table 1, and the error bars in all figures, represent 95% confidence intervals on the linear model's intercept estimate. We calculated these intervals using two methods: the robust variance (or *sandwich*) estimator (results are reported in Table 1 in the main text) and a bootstrapping approach. This approach involves resampling the data set N times and calculating $\hat{C}_{S,P}$ for each data set to estimate the confidence interval. Results can be seen in Table S1.

Signal	Response	$\left(\hat{C}_{S,P} + \delta_{0.025}, \hat{C}_{S,P} + \delta_{0.975}\right)$ (bits)
6. TRAIL	Casp-8 activity	$(1.01 - 0.02,\ 1.01 + 0.02)$
7. TRAIL	Casp-8 activity (live cells)	$(1.01 - 0.01,\ 1.01 + 0.02)$
8. TRAIL	Casp-3 activity	$(0.56 - 0.02,\ 0.56 + 0.02)$
9. Casp-8 activity	Casp-3 activity	$(1.23 - 0.02,\ 1.23 - 0.02)$
11. α-factor	*pFUS1*-GFP	$(2.26 - 0.08,\ 2.26 + 0.09)$
13. Bacterium	neutrophil motion	$(1.82 - 0.12,\ 1.82 + 0.12)$
14. cAMP	*Dictyostelium* motion	$(2.19 - 0.13,\ 2.19 + 0.14)$
15. TRAIL	% dead (HeLa; resampled)	$(2.44 - 0.02,\ 2.44 + 0.02)$
16. TRAIL	% dead (HeLa; FACS)	$(3.41 - 0.03,\ 3.41 + 0.02)$
17. TRAIL	% dead (MCF10A)	$(3.35 - 0.04,\ 3.35 + 0.03)$

Table S1: Select estimates from Table 1 in the main text, but with confidence intervals calculated from bootstrapped data sets ($N \geq 100$, where N is the number of bootstrap samples). δ_p is the difference between the value at the pth percentile and the mean of the bootstrap distribution.

1.1.3 Finding the optimal number of bins

Generating the contingency table relies on a particular discretization or binning of the data. As mentioned above, the signal values used to generate experimental data often represent a natural set of signal bins (e.g. Figure 1C and D of the main text). The number of response bins to generate, however, is not clear *a priori*, and the value of R_B has a large impact on estimates of I. On one extreme, if we set $R_B = 1$, all of the signals will give the same responses, resulting in a mutual information of 0. Alternatively, we could choose a number of bins, R_B, so large that every

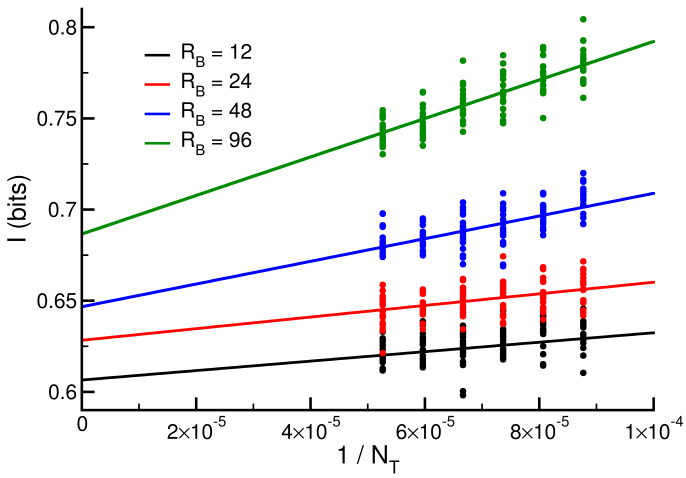

Figure S1: Representative linear models for estimating mutual information at infinite sample size with various numbers of response bins. Here we use experimental data composed of 1000 cells per each of 19 TRAIL concentrations. In this figure we are estimating the mutual information between the level of caspase-8 activity in HeLa cells in response to a uniformly distributed set of TRAIL concentrations. We then calculate the mean mutual information as a function of inverse sample size ($\frac{1}{N_T}$) by taking $n = 20$ independent subsets of the data per sample size. Shown here are the calculated mutual information values for each sampled data set. Calculation of the linear model's intercept provides us with an estimate of mutual information at infinite sample size for a particular number of response bins.

response bin contains exactly one response value in the contingency table:

$$K' = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{cccccccccccc} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 & r_8 & r_9 & r_{10} & r_{11} & r_{12} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

(where again we have used an arbitrary data set as an example). This results in a (spuriously) high mutual information–note that, in this case, if we randomly shuffle the signal value that gives any particular output, we will get the same mutual information:

$$K'_{rand} = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \begin{array}{c} \begin{array}{cccccccccccc} r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 & r_8 & r_9 & r_{10} & r_{11} & r_{12} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Since I generally increases with an increasing R_B (note the increasing intercept for the data in Figure S1), we must find an optimal value of R_B that accurately represents the mutual information in the underlying data without artificially inflating it.

Our approach to solving this problem is broadly inspired by previous approaches, particularly that of Cheong *et al.*, with some slight modifications (3; 5). Each calculation of the mutual information requires some specified number of bins, so we defined both an initial number of bins to use as well as the number of bins by which to increment. For any given R_B value, we generated the bins themselves (*i.e.* the actual range of response values in the data that belongs to each bin) so that the total number of observations $k_{r,j}$ for each response bin is (roughly) equal across all the response bins under the signal distribution given by the data set (3; 5). We then generated the contingency table and estimated I using the linear extrapolation procedure explained above in section 1.1.2.

Plotting I *vs.* the total number of response bins (Figure S2) does indeed demonstrate that mutual information increases essentially monotonically with increasing R_B. For each value of R_B, we also generated N_R contingency tables with some specified sample size with randomly sampled entries. We calculated I for each one of them; the value of I in these randomized data sets also increases with increasing R_B, eventually generating significantly non-zero mutual information where there should be none (Figure S2).

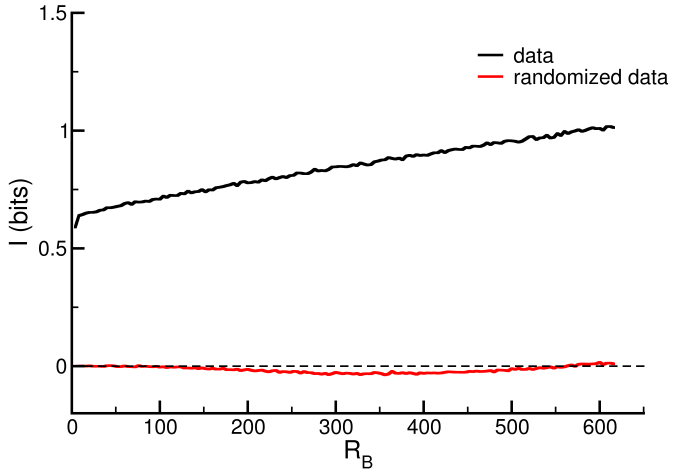

Figure S2: Here we show a representative graph of mutual information (I) as a function of the number of response bins. The data is the same from which Figure S1 is generated and each point shown here is the y-intercept retrieved from the linear extrapolation procedure outlined in section 1.1.2. In red is one randomization of the actual data set, shown in black.

Cheong *et al.* obtained an optimal range of bin numbers for each data set via visual inspection of plots like those in Figure S2 (3). While this is an effective approach, the large number of data sets and variants in our case prevented us from visually analyzing every case. We thus defined a

uniform criterion for choosing the optimal number of bins, defined as the value of R_B that gives the largest value of I, subject to the constraint that the 95% confidence interval from the corresponding randomized data *must include* 0. In other words, we chose an R_B that maximizes the I in the data, but where the randomized data gives mutual information that is not significantly greater than 0.

The range of R_B values that provides this maximum depends on the total number of data points (N_T) and on the amount of information present in the data itself; it is thus difficult to define a uniform range of bin numbers to consider for every data set. Therefore we implemented a method to automate the search for the optimal number of bins. In essence, this method iteratively increments the number of bins used in the calculation until a prespecified number of randomized calculations (in our case, 3) for consecutive increments were biased (*i.e.* significantly above zero).

The discussion above assumes that S_B is fixed at some particular number of signal values, as is typical when generating experimental data. In some of the systems we considered, however, we needed to find an optimal set of signal bins in addition to response bins. This was particularly true of spatial quantities like the angle between the bacterium and the neutrophil (see section 3). In those cases, we adapted this method to identify the optimal number of bins in both signal and response space.

1.2 Channel Capacity

As mentioned in the main text, the channel capacity is the supremum of the mutual information over all possible signal distributions:

$$C = \sup_{p_S(s)} I\left(S; R\right).$$

Estimating the channel capacity thus involves using the estimate of mutual information obtained from the procedure defined in section 1.1 to search the space of signal distributions and find the one that maximizes I. Since the set of such distributions is obviously infinite, an exhaustive search of all well-defined signal distributions is impossible. Following the example of Cheong *et al.* (3), we implemented a grid-based search, limited to a set of unimodal and bimodal Gaussian distributions in addition to a uniform distribution of signal values and the distribution present in the original data set (if it is not already uniform). Additionally our estimate is limited by the finite set of signal values over which the data itself is collected. We thus define a new quantity as the *estimated* channel capacity given the set of sampled signal values S and the set of probability mass functions P attempted in our optimization:

$$\hat{C}_{S,P} \leq C$$

giving us a lower bound on the true channel capacity.

1.2.1 Unimodal signal distributions

We generated a range of unimodal signal distributions of the form:

$$G_U(s) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(s-\mu)^2}{2\sigma^2}}.$$

Since we can sample only a subset of possible signal distributions we limit the potential mean values, μ, to a set of 4 evenly spaced values between the minimum and maximum signal values in the data set (S_{min} and S_{max}, respectively):

$$\mu \in \{f \cdot (S_{max} - S_{min}) + S_{min}\} \mid f \in \{0.2, 0.4, 0.6, 0.8\}\}.$$

For each μ we generate a range of σ values by calculating a maximum standard deviation:

$$\sigma_{max} = \min\left(\frac{\mu - S_{min}}{3}, \frac{S_{max} - \mu}{3}\right).$$

This constrains the signal distributions so that at least 99% of the area under the distribution falls between S_{min} and S_{max} and allows us to use a range of standard deviation values by sampling increasing fractions of σ_{max}:

$$\sigma \in \{f \cdot \sigma_{max} \mid f \in \{0.2, 0.4, 0.6, 0.8, 1\}\}.$$

1.2.2 Bimodal signal distributions

We also implemented a range of bimodal signal distributions of the form:

$$G_B(s) = \frac{w_0}{\sigma_0\sqrt{2\pi}} \mathrm{e}^{-\frac{(s-\mu_0)^2}{2\sigma_0^2}} + \frac{w_1}{\sigma_1\sqrt{2\pi}} \mathrm{e}^{-\frac{(s-\mu_1)^2}{2\sigma_1^2}}$$

where w_0 and w_1 are weighting coefficients such that $w_0 \in \{0.4, 0.5, 0.6\}$ and $w_1 = 1 - w_0$. In order to construct these distributions we first defined a minimum difference between μ_0 and μ_1:

$$\mu_D = \frac{S_{max} - S_{min}}{5}.$$

We used μ_D to construct a series of pairs (μ_0, μ_1) such that

$$\mu_D + S_{min} \leq \mu_0 < S_{max}$$

$$\mu_0 + \mu_D \leq \mu_1 < S_{max}$$

and μ_0 is incremented in steps of μ_D. Similarly to the unimodal signal distributions, both means μ_0 and μ_1 have multiple, evenly spaced standard deviations, σ_0 and σ_1 that are fractions of some maximum standard deviations, $\sigma_{0,max}$ and $\sigma_{1,max}$. These values are constrained so that these distributions have both a local minimum between μ_0 and μ_1 and 99% of their area between S_{min}

and S_{max}:

$$\sigma_{0,max} = \min\left(\frac{\mu_1 - \mu_0}{4}, \frac{\mu_0 - S_{min}}{3}\right)$$

$$\sigma_{1,max} = \min\left(\frac{\mu_1 - \mu_0}{4}, \frac{S_{max} - \mu_1}{3}\right).$$

The individual σ_0 and σ_1 values are then:

$$\sigma_0 = f \cdot \sigma_{0,max}$$

$$\sigma_1 = f \cdot \sigma_{1,max},$$

where $f \in \{0.2, 0.4, 0.6, 0.8, 1\}$.

1.2.3 Weighting the data

With this set of unimodal and bimodal signal distributions, we can determine how the mutual information of a particular data set varies with different signal distributions in order to estimate the channel capacity. To do this, we modified the original contingency table in order to recalculate the mutual information according to each new signal distribution. Each signal bin s_i corresponds to a range of signal values between, say, $s_{i,min}$ and $s_{i,max}$ and yields a corresponding number of observations in the contingency table, k_{ij} for each response bin r_j. For any unimodal or bimodal signal distribution $G_A(s)$, we calculated the new value for this entry in the contingency table $k'_{ij} = \frac{p'(s_i)k_{ij}}{p(s_i)}$ where

$$p'(s_i) = \int_{s_{i,min}}^{s_{i,max}} G_A(s)\, \mathrm{d}s$$

is the new probability of observing some signal value s_i and $p(s_i)$ is the uniform probability of observing that signal bin s_i. We can use this to generate a new contingency table, and calculate the relevant quantities:

$$N'_T = \sum_i^{S_B} \sum_j^{R_B} k'_{ij},\ k'_{s,i} = \sum_j^{R_B} k'_{ij},\ k'_{r,j} = \sum_i^{S_B} k'_{ij}.$$

The procedure produces a new contingency table that has the same number of entries as the original one. For each distribution $G_A(s)$ that we considered, we used the procedures described in section 1.1 to estimate the mutual information for that particular distribution. The maximum mutual information over all signal distributions calculated is our estimate of the channel capacity $\hat{C}_{S,P}$ which we will refer to as simply $\hat{C}$ for notational convenience. As mentioned above, the errors reported for $\hat{C}$ represent the 95% confidence interval for the intercept estimated by the linear extrapolation procedure (section 1.1.2).

2 Additional Experimental Calculations

2.1 Control calculations

As mentioned in the main text, we examined the channel capacities between the activities of the initiator caspase (IC; cleaved caspase 3) and both the effector caspase (EC; cleaved PARP) and terminal cellular phenotype. We found that the IC to EC channel capacity exceeded 1.2 bits. This confirms the nature of IC as an intermediate component in the TRAIL signaling network due to the relative increase in information when using IC as the input distribution to the channel capacity estimation instead of TRAIL.

2.2 Population size dependence of single-cell channel capacity estimation

Given our large data set, we investigated how $\hat{C}$ would vary for individual cells as a function of population size; as mentioned in the main text, the population-level $\hat{C}$ has a clear dependence on the size of the population. We expected to find that as the sample size increases the estimators describing the response distribution will be sufficiently accurate to prevent the need to calculate $\hat{C}$ from the entire data set of over 1.2 million cells (which is computationally expensive). We confirmed this empirically, upon estimation of the single-cell $\hat{C}$ for increasing subsets of our FACS-generated data, using sample sizes of 500, 1000, and 2000 cells per TRAIL concentration.

2.3 Dose-dependent scaling

In our data we observed that IC activity levels were substantially higher in dead than live cells, most likely due to the variety of positive feedback mechanisms present in the caspase cascade (6). Because this additional cleavage of caspase 3 in dead cells occurs downstream of the cell's commitment to apoptosis, it could be considered a consequence of the cell's phenotypic outcome rather than as an intermediate factor contributing to it. Since the channel capacity estimation is time-dependent (and capturing the exact moment of cell death for every cell is technologically infeasible), we proceeded to examine the impact that post-commitment IC activity has on our estimates for channel capacity between TRAIL dose and IC activity level. We therefore performed our analysis separately for live and dead cells by partitioning them into these two groups according to the threshold effector caspase response, t_{EC} (7); we calculated this quantity by estimating the minimum density between the two peaks of the bimodal EC activity distribution: $\log_{10}(t_{EC}) = 2.85 \pm 0.05$ (8). We then plotted the dose response data to determine how response varies with TRAIL concentration. These plots show clearly that only initiator caspase activity scales with TRAIL dose, and it does so only among living cells (Figure S3). The value of $\hat{C}$ between TRAIL and IC activity in living cells is approximately 1.01 bits as shown in Table 1 in the main text, essntially the same as the estimated channel capacity between TRAIL and IC activity in all cells. Mean effector caspase activity in living and dead cells in addition to mean initiator caspase activity in dead cells does not significantly vary for differing doses of TRAIL (Figures S3 and S4) and as a result, we did not estimate the channel capacity for these dose-response relationships.

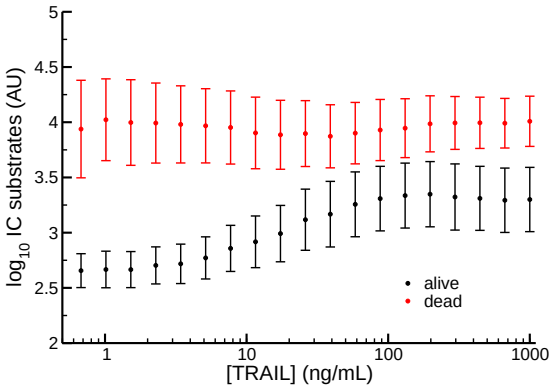

Figure S3: Initiator caspase activity scales with TRAIL among living cells. Shown for each TRAIL concentration are the sample mean and standard deviation ($n \approx 60,000$ cells)

2.4 Resampling experimental data

In order to estimate the channel capacity for the pheromone signaling network in yeast, we reconstructed dose-response data shown by Bashor *et al.* in their Figure 4D (9). In this case the signal distribution was a set of logarithmically spaced α-factor concentrations and the corresponding response was *pFUS1*-GFP fluorescence. With this data we constructed a series of dose-dependent Gaussian distributions defined by the mean and standard deviation of the *pFUS1*-GFP response given some α-factor concentration. From these distributions we sampled 100 values for each of 10 α-factor concentrations in order to construct a dose-response data set from which we could estimate the channel capacity (Figure S5). We similarly performed this procedure for estimating the population-level channel capacity for the set of MCF10A and HeLa cells shown in Figure 3B of the main text. In this case, the mean and standard deviation for a particular TRAIL dose refer to the number of living cells in a given population.

3 Spatial Channel Capacity Estimation

In order to estimate the spatial channel capacity between a motile cell undergoing chemotaxis and its target that is producing some chemical gradient, we constructed signal/response pairs from angles between the cell and its target. We used the CellTrack program developed by Sacan *et al.* (10) to output text files containing frame-by-frame coordinates for the edges of the cells and their centers of mass (COM) as shown in Figure S6A. We used these coordinates to calculate time-dependent signal and response angles (Figure S6B).

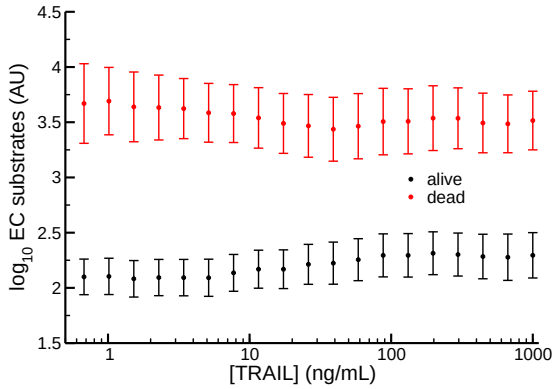

Figure S4: Effector caspase activity is invariant with TRAIL among both living and dead cells. The data set is identical to that in Figure S3.

3.1 Neutrophil motion

We initially analyzed the motion of a neutrophil that is "chasing" a bacterium from a classic movie taken in the 1950s (www.youtube.com/watch?v=Kb-m1uDoWfU). For the purposes of our calculation, we assume that the neutrophil is in fact following a chemical gradient generated by the bacterium. In this case, the signal corresponds to the angle, termed θ_1, between the bacterium at a particular frame in the movie, time t, and the neutrophil at another time $t + \Delta t_1$. The subsequent response angle, θ_2, is that of neutrophil motion between time $t + \Delta t_1$ and time $t + \Delta t_1 + \Delta t_2$. These angles then comprise the signal and response distributions used to calculate $\hat{C}$ of the system. A visual representation of this calculation can be seen in the main text (Figure S6A&B).

We calculated the signal and response angles between the neutrophil COM and bacterial COM relative to the x-axis unit vector in the Cartesian coordinate system. This method is similar to one outlined by Burov *et al.* derived to provide more directional information than mean squared displacement for analysis of random walks (11). We employ a "windowed" data collection method; given some starting time, t, we calculate an arbitrary signal and response angle pair, requiring information from time points, $t + \Delta t_1$ and $t + \Delta t_1 + \Delta t_2$. In our windowed data collection, the next pair of angles is calculated using t incremented by one frame: $t = t + 1$. To confirm that the estimated channel capacity was not an artifact of the chosen time delay values, Δt_1 and Δt_2, we explored the nearby $(\Delta t_1, \Delta t_2)$-space and discovered that $\hat{C}$ is relatively robust to Δt_1 and Δt_2 as seen in Figure S6C.

3.2 Dictyostelium motion

The next movie we analyzed is that of a *Dictyostelium* cell following a cAMP gradient (Movie S1, provided courtesy of Susan Lee and Richard Firtel, Section of Cell and Developmental Biology,

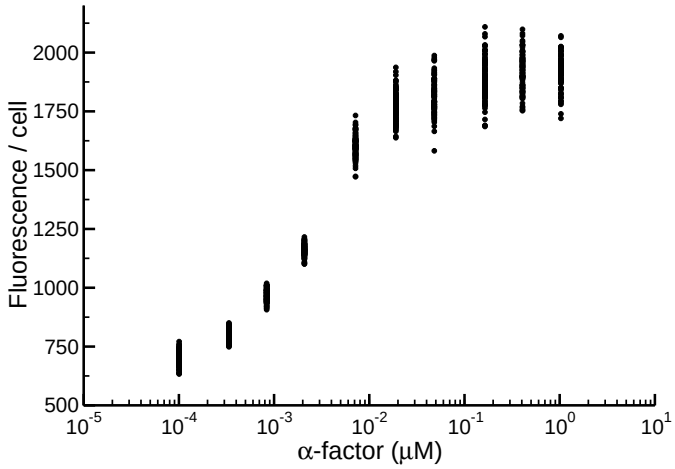

Figure S5: Resampled data from Figure 4D in Bashor *et al.* (9).

Division of Biological Sciences, University of California, San Diego, La Jolla, CA). In this movie, *Dictyostelium* responds to cAMP introduced by a pipette tip which changes location periodically. Since the pipette tip remains stationary between location shifts, we can employ our original calculation used for the neutrophil/bacterium data and omit the Δt_1 parameter (the largest value of $\hat{C}$ occurs when $\Delta t_2 = 18$). This omission is valid since θ_1 is identical for a range of θ_2 values (*i.e.* no motion in the gradient source/change in the signal).

4 Simple Model

As mentioned in the main text, our initial model takes the form:

$$R = (R_{max} - R_{min}) \cdot \frac{S^n}{S^n + K^n} + R_{min} + \epsilon \tag{2}$$

where the normally-distributed noise term $\epsilon \sim N(0, \sigma)$ has some chosen standard deviation, σ. The parameter values chosen for the base model (shown in Figure 3) are as described in the Materials and Methods section of the main text: $K = 10$, $n = 6$, $R_{max} = 30$, and $R_{min} = 20$. For all models discussed in the paper, the response threshold governing an individual cell's fate is positioned such that half of the signal values produce mean responses below the threshold and half produce mean responses above the threshold.

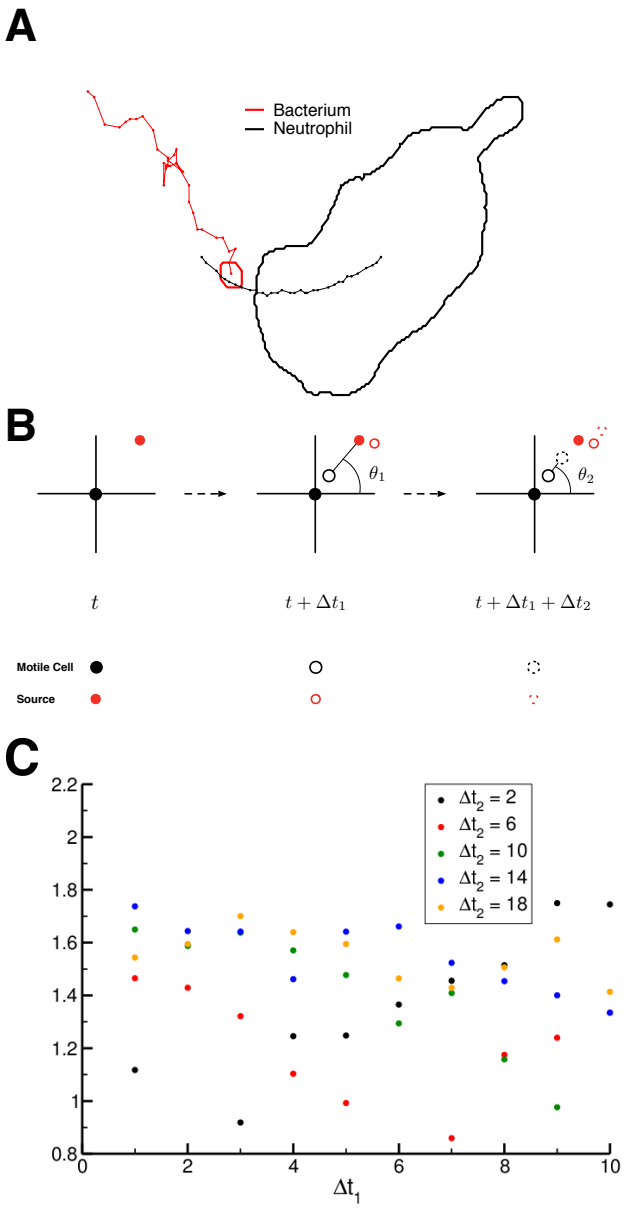

Figure S6: **A**. A representative trajectory from the neutrophil movie. Points are centers of mass for the bacterium (signal source, red) and neutrophil (motile cell, black). The bold, outlined areas show the cells? perimeters in the trajectory's first frame. Clearly seen here is the bacterium's stochastic random walk-like motion and the neutrophil's smoother tracking of the resulting gradient. **B**. Black circles represent the motile cell and red circles represent the signal source (either a micropipette or bacterium). Since cells do not instantaneously detect or respond to extracellular stimuli, filled, solid, and dashed circles represent the motile cell and signal source at initial (t), signal-delayed ($t + \Delta t_1$), and response-delayed ($t + \Delta t_1 + \Delta t_2$) time, respectively. We can then calculate the mutual information between the signal (θ_1) and response (θ_2) angles. **C**. Spatial $\hat{C}$ for the neutrophil as it depends on Δt_1 and Δt_2 for select values.

4.1 Choosing signal values

For our initial analysis of the model, we selected evenly-spaced signal values to achieve responses 10% above the minimum response and 10% below the maximum response (the *transition zone*):

$$0.1 \cdot (R_{max} - R_{min}) + R_{min} \leq R \leq 0.9 \cdot (R_{max} - R_{min}) + R_{min}$$

Through simple algebra we show that the resulting minimum and maximum signal values, (S_{min} and S_{max}, respectively) are:

$$S_{min} = K \cdot \sqrt[n]{\frac{1}{9}}$$

and

$$S_{max} = K \cdot \sqrt[n]{9}$$

This prevents selection of signal values that would produce extremely high or low responses, since sampling more of these responses relative to intermediate responses would reduce the estimated channel capacity.

4.2 Varying n

In order to determine the effect of our chosen $n = 6$ on this model's estimated channel capacity (both single-cell and population-level), we varied n between 2 and 10. We see in general from Figures S7 and S8 that this variation produces minimal difference between models; qualitatively, models with different n are nearly identical.

4.3 Estimated channel capacity saturation with population size

As discussed in the main text, the fraction of a group of cells making a particular signal-dependent decision is the statistic used to determine collective response for the population-level estimation of the channel capacity. By calculating this statistic over a number of replications, we effectively construct a sampling distribution for the fractional response to some arbitrary signal. Since the standard deviation of this distribution (*i.e.* the standard error of the statistic) is dependent on sample size, we observe an inverse correlation between the size of the population and the standard error of the fractional response (Figure S9). In other words, the larger the population, the lower the fluctuations in the percentage of cells that respond to some signal, and the higher the population-level channel capacity becomes (see Figures 3B and D in the main text). In our calculations for the Gaussian model (Equation 2), the estimated channel capacity eventually saturates at some level as population size increases. We should note that this is a simple consequence of the fact that we limit our sampling in this model to a particular number N_S of singal values the increasing regime of the *single-cell* dose-response curve. As a result, when noise is high enough, the population-level channel capacity approaches its theoretical maximum, which is the entropy of a uniform distribution over the sampled signal values ($\hat{C} \sim \log_2(N_S)$ bits). Changing how the signal values are sampled thus has a strong influence on estimates of the population-level channel capacity, an issue that is discussed in detail in sections 4.5 and 7.3 below.

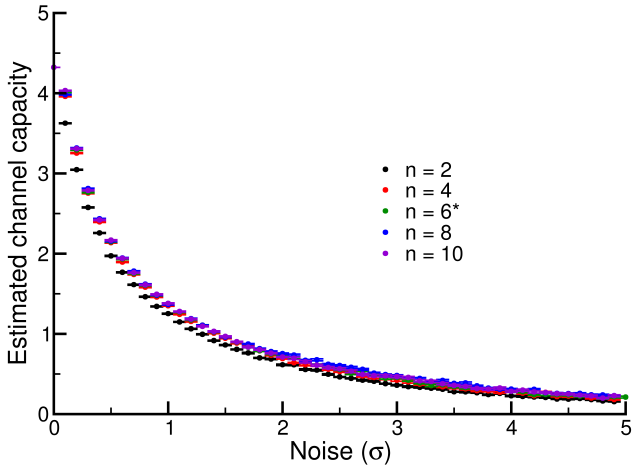

Figure S7: Single-cell estimated channel capacities ($\hat{C}$) with respect to noise for a range of n values. The starred $n = 6$ denotes the value used in all other calculations based on this model. There is minimal difference between models where $n > 2$ and even the model with $n = 2$ displays qualitatively similar behavior to the others. Error bars denote 95% confidence about the intercept estimate (see section 1.1.2)

4.4 Maximal fractional response

At high levels of noise, we observe another important feature of the population response: the inability to effect a universal population response at arbitrarily high signal levels. In other words, no matter how much signal is present in the environment, there will still be a fraction of cells in a population that does not respond. This is plainly observed graphically in Figure S10; even as the response saturates at low and high signal levels, there is sufficient noise such that a subpopulation of cells at these signal levels fall above and below the threshold, respectively. The corresponding population dose-response curve thus exhibits saturating, incomplete responses both at low and high signal levels (Figure S11).

As mentioned above, we chose the threshold in the response (*i.e.* the value of R above which an individual cell decides to die or undergo some other cell-fate decision) so that this saturation is *symmetric*. In other words, if 25% of the cells survive at saturating levels of signal, then 25% of them will die even when there is essentially no signal at all in our model (Figure S10). Note that, in our experimental data for HeLa cells, we found that 100% of cells survive when the population is not treated with TRAIL, and around 50% of them die at saturating TRAIL levels (Figure 2A in the main text). In our model, this would correspond to moving the threshold to higher values, so that no cells respond at low signal, and only half respond at high signal. Since I, and thus $\hat{C}$ are invariant to this kind of transformation (*i.e.* changing the absolute range of the response value), we focused on a symmetric threshold for simplicity, without loss of generality.

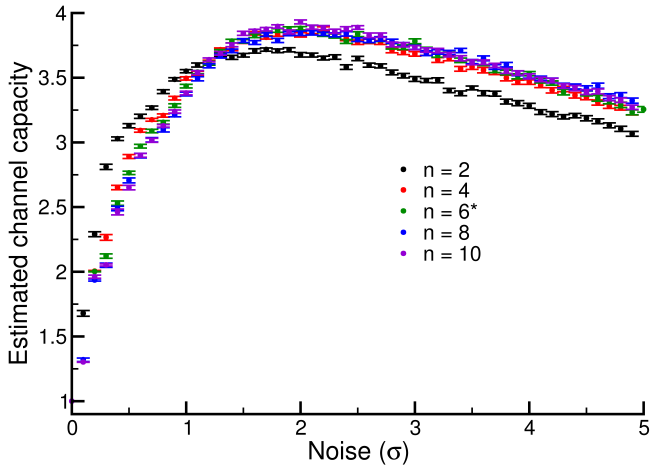

Figure S8: Population-level $\hat{C}$ with respect to noise for a range of n values. Again we see little difference between models with different n with the minor exception of $n = 2$. We do observe a slight shift in the amount of noise producing maximal $\hat{C}$, but the qualitative trends are essentially identical. Error bars are as in Figure S7.

4.5 Population $\hat{C}$ dependence on signal spacing and noise

As mentioned in the main text, we sampled a fixed number of signal values corresponding to responses between 10% and 90% of the maximal *population* response for a given noise value. Upon reaching a noise threshold, the population-level $\hat{C}$ is effectively constant as noise approaches 0 (Figure S12), contradicting the existence of a nontrivial level of noise producing maximal population-level $\hat{C}$ shown in the main text (Figure 3C). We also observed that by simply increasing the number of signal values sampled in the single-cell transition region decreases the amount of noise required to reach the maximal population-level $\hat{C}$ (Figure S13). Both of these strategies for increasing (or maintaining) population-level $\hat{C}$ require increasingly small ΔS. Since many *in vivo* inducers of intracellular signaling are cytokines that are themselves subject to fluctuations in concentration, we posited that this variability in the signal would prevent optimal population-level $\hat{C}$ at arbitrarily low noise.

To confirm this numerically, we introduced another noise term into the simple model (the *signal* noise, ϵ_s, as opposed to the original *response* noise, ϵ) governing the limit of signal accuracy for our populations of simulated cells. This alternate form of the model (modified from Equation 2) has the following form:

$$R = (R_{max} - R_{min}) \cdot \frac{(S + \epsilon_s)^n}{(S + \epsilon_s)^n + K^n} + R_{min} + \epsilon$$

where the signal noise term ϵ_s is normally distributed and is sampled independently for each *population* of cells: $\epsilon_s \sim N(0, \sigma)$. This procedure simulates stochastic cytokine production and

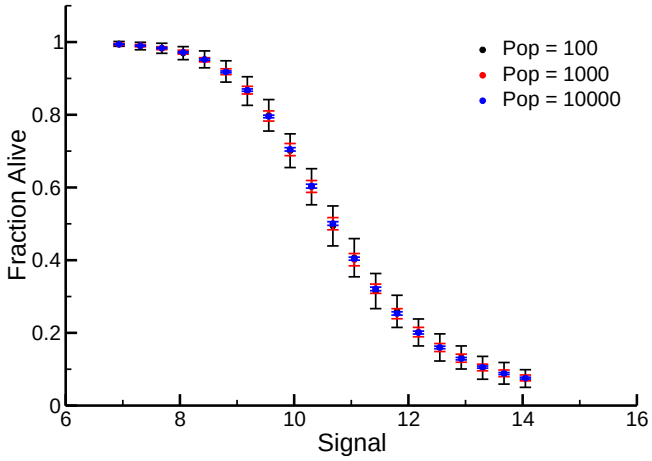

Figure S9: Population-level dose-response curves for multiple population sizes

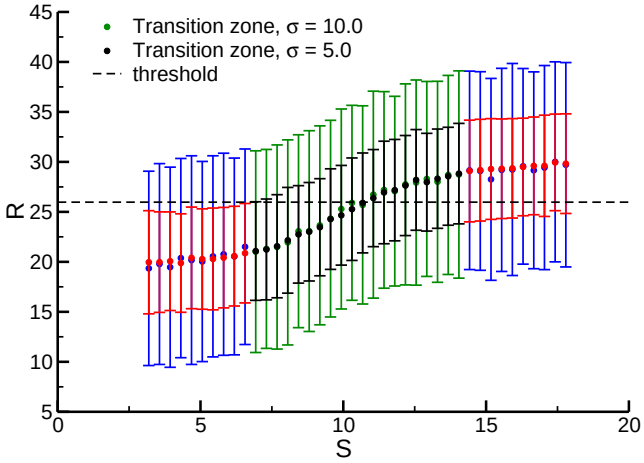

Figure S10: Single cell response curve at high noise values

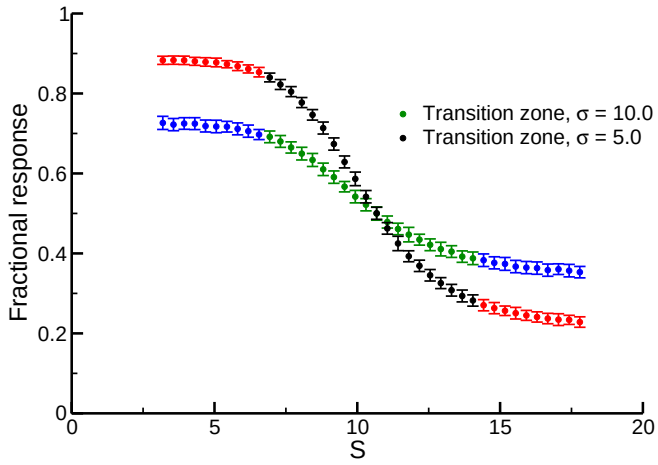

Figure S11: Population response curve at high noise values (population size = 1000)

distribution to cellular populations, and it is distinct from the application of the original noise term, ϵ, that was applied to each individual cell. The population-level $\hat{C}$ can been seen in Figure 3D in the main text for select values of signal noise, revealing that even in the presence of very small signal noise (1% of the signal needed to achieve a half-maximal response), there exists a nontrivial level of noise that maximizes information available to cellular populations.

The question then becomes: what is a reasonable level of signal noise? Unfortunately, *in vivo* distributions for cytokines like TRAIL have not been measured in sufficient detail for us to characterize the typical size of such fluctuations. To estimate this, we considered a prototypical signaling environment: namely, the intestinal crypt in mammals. This cylindrical structure contains about 1000 cells that line a fluid-filled interior lumen with a volume of about 5×10^{-12} L (12; 13). As can be seen from Figures 2A and 2B in the main text, the population-level transition in cell death occurs from about 1 to $10^2 - 10^3$ ng/mL. This would translate to a range of $\sim 10^2 - 10^4$ or 10^5 molecules of TRAIL in the lumen of the intestinal crypt.

Since cytokines like TRAIL are actively synthesized by cells, and would be lost from the lumen either due to active degradation or diffusion away from the crypt, those cytokines would, at the very least, be subject to standard Poisson-style birth-death fluctuations. The relative error in signal values (*i.e.* the coefficient of variation, or $\sigma_s/\langle s \rangle$) would be $1/\sqrt{m}$, where m is the number of molecules (14; 15; 16; 17). In the case of the intestinal crypt described above, we would expect around 10^4 molecules, corresponding to a relative error of $\sim 1\%$.

We should note that the estimate of $\sim 1 - 0.1\%$ signal error ignores any "extrinsic" sources of noise in the signal (15; 16; 17), and as such likely represents an underestimate of signal fluctuations in biological systems. Interestingly, we observed relatively large dynamic ranges for the population-level responses in both HeLa and MCF10A cells (over an order of magnitude of TRAIL concentrations, see Figure 2 of the main text), which may indicate that signal noise has

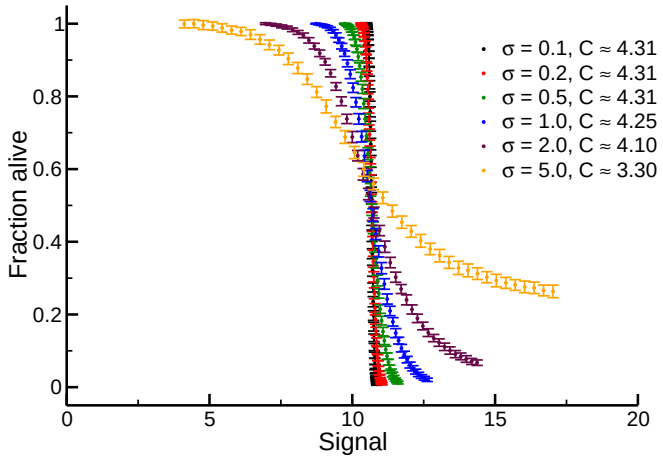

Figure S12: By sampling evenly-spaced signal values in the population dose-response transition zone (which varies with the level of noise in individual cells), we observe that the population-level $\hat{C}$ becomes invariant to noise levels as below a certain threshold.

played a role in shaping population-level response. In any case, although very small levels of single cell noise could theoretically produce high population-level $\hat{C}$ (Figure S12), the presence of signal variability in biologically realistic systems requires higher levels of noise, so that signal values can be spaced at reasonable distances from one another while maintaining low variability in the population-level response.

5 Analytical results for Gaussian responses to signal

Numerical calculations on our original signal-response model suggested a number of interesting underlying relationships between noise and single-cell/population-level channel capacity. In order to characterize those relationships more fully, we developed an analytical framework for the model. Our original version utilized a discrete set S of signal points, and as such represents a type of "Gaussian mixture model" commonly encountered in communication theory and other areas of computer science. Unfortunately, there is currently no closed-form solution for the entropy of a Gaussian mixture model (18). In order to make analytical progress on the model, we created a closely related version where S is now construed to be a *continuous*, rather than discrete, random variable. Though some parameters differ in this version (*i.e.* the minimum and maximum mean response), the core structure of the model, including the signal-response relationship and the variability of single-cell response given a particular signal value, remains the same.

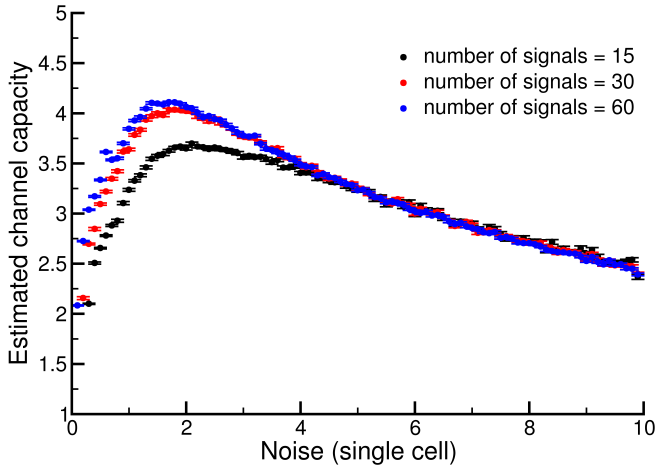

Figure S13: Increasing the number of sampled signal values in the transition zone of the *individual* cell's dose-response curve (*i.e.* independently of the population dose-response transition zone) results in a decrease in optimal noise level and an increase in estimated channel capacity.

5.1 Model definition

Our analytical calculations of the continuous, Gaussian signal-response model are based on the following definitions:

1. R is the response variable
2. S is the signal variable
3. $p\left(R|S\right)$ is normally distributed:

$$p\left(R|S=s\right)=\mathrm{N}\left(\mu\left(s\right),\sigma\right)$$

   with some arbitrary σ and a signal-dependent $\mu\left(s\right)$.

4. In our original signal-response model, we had $\mu\left(s\right)=r_{min}+\left(r_{max}-r_{min}\right)\dfrac{s^n}{s^n+K^n}$, where K is the signal value giving a half-maximal response and r_{min} and r_{max} are the minimum and maximum average responses, respectively. To simplify the calculation, in this model we define $r_{min}=0$ and $r_{max}=1$ so that $p\left(R|S=s\right)=\mathrm{N}\left(\dfrac{s^n}{s^n+K^n},\sigma\right)$. Since the units on R are arbitrary and don't influence the calculation of $I(S;R)$, we can make this simplification without loss of generality.

5.1.1 Calculating entropies and channel capacities

Our initial goal with this model is to characterize the mutual information of this signal-response relationship. We know from the definition of mutual information that:

$$I\left(S;R\right)=h\left(R\right)-h\left(R|S\right)$$

where $h\left(R\right)$ denotes differential entropy and is defined as:

$$h\left(R\right)=-\int_{-\infty}^{\infty}p\left(r\right)\log_2 p\left(r\right)\mathrm{d}r,$$

and the base of the logarithm denotes the units of information. Here we use base 2, resulting in units of bits. The conditional differential entropy is then:

$$h\left(R|S\right)=-\int p\left(s\right)\int p\left(r|s\right)\log_2 p\left(r|s\right)\mathrm{d}r\mathrm{d}s.$$

Note that we have dropped the bounds on the integral for notational convenience; the domain $(-\infty,\infty)$ is implied for both S and R. In order to calculate this, and related quantities, it is helpful to know the probability density of signal $p(s)$. To calculate C, we need to find the density that maximizes $I(S;R)$. Laughlin first introduced the principle of "gain adaptation" to show that I is maximized when the input (signal) distribution is the first derivative of the mean signal-response curve (as can be seen in Figure S14) (19). In other words, the optimal signal distribution is obtained when considering $\mu(s)$ to be the Cumulative Distribution Function (CDF) of the signal. We can thus calculate the probability density of signal as:

$$p\left(s\right)=\frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{s^n}{s^n+K^n}\right)=\frac{s^n n}{s\left(s^n+K^n\right)}-\frac{n\left(s^{2n}\right)}{s\left(s^n+K^n\right)^2}$$

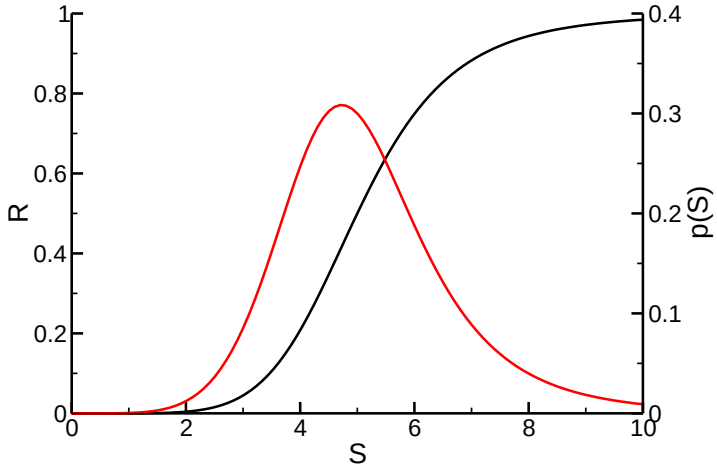

Figure S14: Calculating the optimal signal probability density for the continuous Gaussian signal-response model. In this case, we have taken $K = 5$ and $n = 6$ in the model, for the purposes of display. Shown in black is the mean response to signal and shown in red is the corresponding probability density that maximizes information transmission.

We have $p(r|s)$ by definition so we can then find both $h(R|S)$ and $p(r)$. We will focus first on $h(R|S)$:

$$\begin{aligned} h(R|S) &= -\int p(s) \int \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{\left(r - \frac{s^n}{s^n + K^n}\right)^2}{2\sigma^2}} \log_2 \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{\left(r - \frac{s^n}{s^n + K^n}\right)^2}{2\sigma^2}} \mathrm{d}r\mathrm{d}s \\ &= \int p(s) \cdot \frac{1}{2} \log_2 \left(2\pi \mathrm{e}\sigma^2\right) \mathrm{d}s \\ &= \frac{1}{2} \log_2 \left(2\pi \mathrm{e}\sigma^2\right) \cdot \int p(s) \, \mathrm{d}s \\ &= \frac{1}{2} \log_2 \left(2\pi \mathrm{e}\sigma^2\right) \end{aligned}$$

which is obviously just the entropy of the conditional Gaussian distribution. We then calculate

the marginal probability density of signal, $p(r)$:

$$
\begin{aligned}
p(r) &= \int p(s) \cdot p(r|s)\, \mathrm{d}s \\
&= \int \left(\frac{s^n n}{s\left(s^n + K^n\right)} - \frac{n\left(s^{2n}\right)}{s\left(s^n + K^n\right)^2} \right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\frac{\left(r - \frac{s^n}{s^n + K^n}\right)^2}{2\sigma^2}} \mathrm{d}s \\
&= \frac{1}{2} \left(\mathrm{erf}\left[\frac{r\sqrt{2}}{2\sigma}\right] - \mathrm{erf}\left[\frac{(r-1)\sqrt{2}}{2\sigma}\right] \right)
\end{aligned} \tag{3}
$$

where erf is the standard error function. Note that this integral was computed using Mathematica (20). This gives us the following integral for the marginal entropy of R:

$$
h(R) = \int \frac{1}{2} \left(\mathrm{erf}\left[\frac{r\sqrt{2}}{2\sigma}\right] - \mathrm{erf}\left[\frac{(r-1)\sqrt{2}}{2\sigma}\right] \right) \log_2 \left(\frac{1}{2} \left(\mathrm{erf}\left[\frac{r\sqrt{2}}{2\sigma}\right] - \mathrm{erf}\left[\frac{(r-1)\sqrt{2}}{2\sigma}\right] \right) \right)
$$

Since the difference of error functions enters the integral under the logarithm, we were unable to compute a closed-form expression for $h(R)$, and by extension $C(S;R)$, in this case. We consider a simplified and more analytically tractable version of this model below (Section 6). We can, however, calculate the marginal entropy of R in this model by using numerical integration. Using standard integration packages in Python (21), we calculated $h(R)$ and $C(S;R)$ and found that the single-cell channel capacity in this model does indeed approach 0 as noise increases (Figure S15).

To better understand this finding, we can plot the two error functions whose difference makes up $p(r)$ (see Figure S15A). Note that, when σ is small, the error functions are essentially two step functions, one that occurs at 0 and the other that occurs at 1 (Figure S15A, red curve). The marginal probability density of R, which is just the difference between these two functions (Equation 3), is thus essentially uniform in this regime (Figure S15B, red curve). When σ is large, however, the error functions undergo a more gradual transition, and the resulting $p(r)$ adopts an apparently Gaussian distribution with mean 1/2 and a standard deviation similar to that of the underlying $p(r|s)$ (Figures S15A and B, black curves). In other words, it appears that $p(r) \to p(r|s = K)$ as $\sigma \to \infty$. To consider this possibility in greater detail, we numerically calculated the Kullback-Leibler divergence between $p(r|s = K)$ and $p(r)$, and found that this divergence rapidly approaches 0 as σ increases (Figure S15D). This indicates that the marginal distribution of the response is limiting to the conditional distribution of the response; since $h(R) \to h(R|S)$ as $\sigma \to \infty$, we have $C(S;R) \to 0$ (Figure S15C).

6 Analytical model: Uniform signal and response distributions (single cell)

The difficulty with calculating $h(R)$ for single cells in the Gaussian signal-response model prevented us from making any further progress on the trade-off between single-cell and population-level information. To overcome this problem, we considered a related model that is based on uniformly-distributed noise. To simplify the integrals and the resulting expressions, the

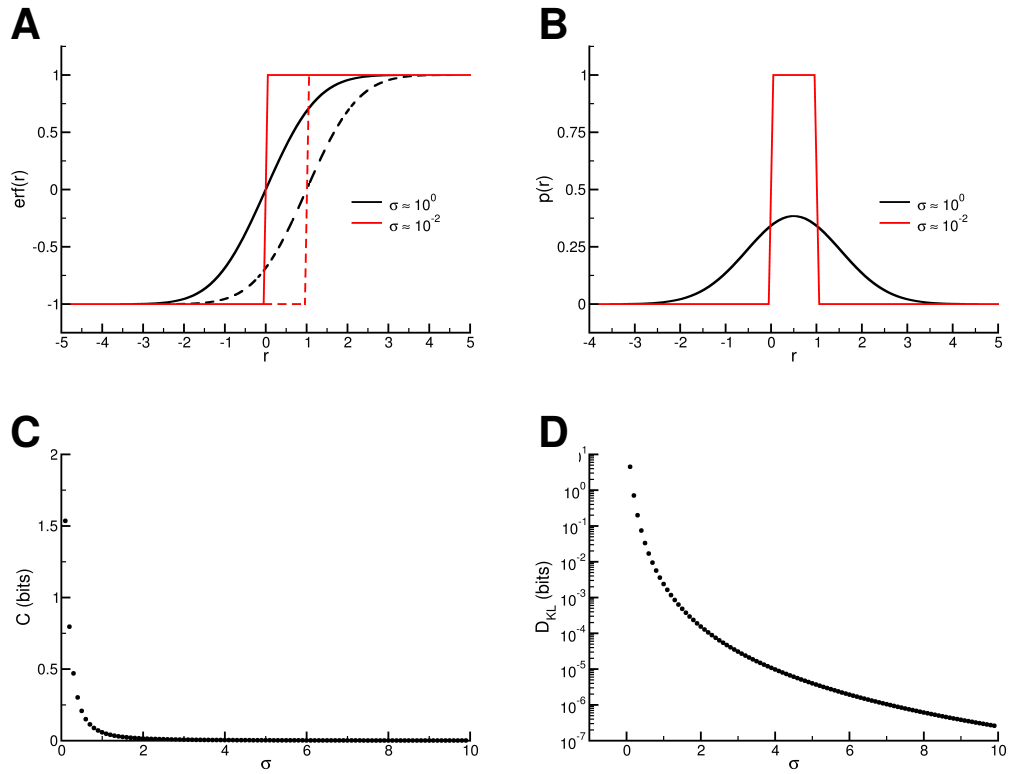

Figure S15: **A**) The marginal probability of r is the difference between two error functions. Here this difference is visualized with high (black) and low (red) σ values. **B**) The marginal probability of r, $p(r)$. As σ approaches 0, $p(r)$ becomes a uniform distribution between 0 and 1, with an entropy of 0 bits. As σ increases, the distribution seems to adopt a Gaussian character with mean 1/2. **C**) Channel capacity resulting from numerical integration of the Gaussian model. The x-axis is the standard deviation, σ, of $p(r|s)$. **D**) The Kullback-Leibler divergence between $p(r|s=K)$ and $p(r)$, $D_{\mathrm{KL}}(p(r|s=K)\,||\,p(r))$. As $\sigma \to \infty$, $D_{\mathrm{KL}}(p(r|s=K)\,||\,p(r)) \to 0$ confirming that $p(r)$ becomes a Gaussian distribution with a standard deviation equal to that of $p(r|s=K)$.

logarithms shown in the following equations are the natural logarithm, and as a result they report the mutual information in units of "nats." Conversion to the more standard units of bits can be achieved simply by multiplying each expression by the factor $\log_2 \mathrm{e}$. To ease comparison with previous figures, this conversion was performed for all of the plots in this section, as well as in Section 7.

6.1 Definitions

With this model, the notation generally remains the same, and the structural differences are outlined in the following definitions. Most notably, the signal-response relationship is linear, and the probability distribution of responses given a particular signal are uniform.

1. The definition of the model in this case is:

$$R = \frac{S}{b} + \epsilon; \quad \epsilon \sim U(-a, a)$$

   where U denotes a uniform distribution. In this case, the mean response at some signal s is just s/b. Note that the parameter a in this model sets the level of "noise" in the conditional response $p(r|s)$, and so a plays a role similar to that of σ in our gaussian models.

2. To ensure that the mean response is restricted to the interval $[0, 1]$, we define the domain of S as $[0, b]$. As a result, we have that all responses lie on the interval $[-a, 1 + a]$.

While this model is much simpler than the Gaussian models analyzed above, it captures the majority of the salient features of those models: the average response is an increasing function of s, and there is some distribution of responses about that mean, the width of which is controlled by a "noise" parameter. As a consequence of these definitions, we have that:

$$p(r|s) = \frac{1}{2a}$$

which is just the uniform probability density on the interval $[s/b - a, s/b + a]$. As with the Gaussian models above, we also need to have some marginal distribution of signal $p(s)$ to calculate relevant information-theoretic quantities. Using again the result of Laughlin, we have:

$$p(s) = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{s}{b}\right) = \frac{1}{b}.$$

Note that we will also define the half-maximal response to be K for consistency in notation; in this case, we simply have $K = b/2$.

In analyzing this model. it is helpful to first define two distinct scenarios depending on the range of values that a can take:

1. $a < \frac{1}{2}$
2. $a \geq \frac{1}{2}$

As can be seen below, these two scenarios set the different bounds of integration one needs to use in order to calculate the marginal distribution of responses $p(r)$. The reason for this is the fact that, when $a < 0.5$ we can always find two signal values $s_1, s_2 \in [0, b], s_1 < s_2$ such that $s_1/b + a < s_2/b + a$. In other words, it is always possible to find some finite regions of signal space whose response distributions *do not* overlap. When $a = 0.5$, or course, we have that the two extreme values of s have probability distributions that *do* overlap, since $0/b + 1/2 = b/b - 1/2$. For $a < 0.5$, the size of this region of complete overlap increases with increasing a.

6.2 Calculating $C(S; R)$ when $a < 1/2$

We will first examine the cases in which not all signal-dependent distributions overlap.

6.2.1 Calculating $p(r)$

The first step in calculating the mutual information of this model is to find the marginal and conditional distributions of one of the variables and since we have the conditional distribution of R by definition, we will first focus on determining the marginal distribution of R.

For any given value of r there are a range of signal values from which it can arise. We define the lower and upper bounds of this range as follows:

- $S_m(r)$ is the minimum signal that can produce r
- $S_M(r)$ is the maximum signal that can produce r

Since $a < 1/2$, there are three distinct regions for r based on the fact that r values within these regions share the functional form of the limits S_m and S_M:

1. $-a \leq r \leq a$
2. $a < r < 1 - a$
3. $1 - a \leq r \leq 1 + a$.

These regions can be visualized in Figure S16. We can use these bounds to construct an analytical expression to the probability density $p(r)$ in some region i of R values:

$$\begin{aligned} p_i(r) &= \int_{S_{m,i}(r)}^{S_{M,i}(r)} p(r|s)\, p(s)\, \mathrm{d}s \\ &= \int_{S_{m,i}(r)}^{S_{M,i}(r)} \frac{1}{2ab} \mathrm{d}s \end{aligned}$$

6.2.2 Region 1

This region is characterized by the fact that $p(r|s = 0) \neq 0 \forall r \in [-a, a]$; in other words, this is the only set of r values that will have any contribution from $s = 0$, which is the minimum possible S value in this model. So we have that $S_{m,1} = 0$ is the lower bound of the integral $p_1(r)$. The upper bound is the largest value of s that allows sampling of a particular r value, which in this case depends on r itself. To find this limit, we note that, if $s/b - a > r$ then $p(r|s) = 0$, so naturally $S_{M,1} = (r + a)b$. To summarize, our bounds for this region are:

$$S_{m,1}(r) = 0$$

$$S_{M,1}(r) = (r + a)\, b.$$

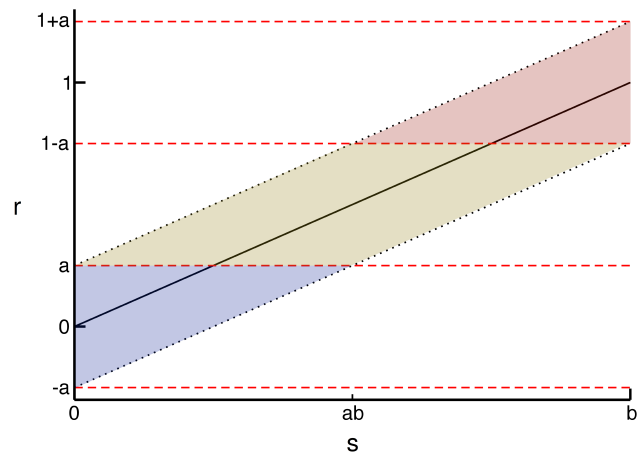

Figure S16: This plot shows a representation of a possible dose-response curve for the uniform model where $a \leq 1/2$; in this case, $a = 1/2$ for simplicity. The solid black line is the average response for some signal-dependent response distribution, and the dotted lines indicate the minimum and maximum responses for the distribution. Red dashed lines delimit distinct response regions; region 1 is blue, region 2 is yellow and region 3 is red. Note that the signal value ab corresponds to the maximum signal value for any response in region 1 and the the value $b - ab$ is the minimum for any response in region 3.

We can then construct the final expression for the probability density for any given r in region 1:

$$\begin{aligned} p_1(r) &= \int_0^{(r+a)b} \frac{1}{2ab} \mathrm{d}s = \frac{1}{2ab}(r+a)\,b \\ &= \frac{r}{2a} + \frac{1}{2}. \end{aligned}$$

6.2.3 Region 2

This region's upper bound is the same as region 1, however the lower bound is the smallest value s that allows sampling of that particular r value. Using reasoning similar to that above, we have:

$$S_{m,2}(r) = (r-a)\,b$$

$$S_{M,2}(r) = (r+a)\,b.$$

The probability of R in region 2 is then:

$$\begin{aligned} p_2(r) &= \int_{(r-a)b}^{(r+a)b} \frac{1}{2ab} \mathrm{d}s = \frac{1}{2ab}\left((r+a)-(r-a)\right)b \\ &= \frac{r+a}{2a} - \frac{r-a}{2a} = \frac{2a}{2a} \\ &= 1. \end{aligned}$$

6.2.4 Region 3

This region is analogous to region 1, since one bound is constant and the other is variable. However in this case the upper bound is fixed at b since all r values in this region have $p(r|s=b) \neq 0$ and b is the largest possible value of s. The lower bound is identical to that in region 2, so we have:

$$S_m(r) = (r-a)\,b$$

$$S_M(r) = b.$$

The probability density for r in region 3 is then:

$$\begin{aligned} p_3(r) &= \int_{(r-a)b}^{b} \frac{1}{2ab} \mathrm{d}s = \frac{1}{2ab}\left(b-(r-a)\,b\right) \\ &= \frac{1-r}{2a} + \frac{1}{2}. \end{aligned}$$

6.2.5 Probability density of the response

The overall probability density $p(r)$ is given by the following piecewise function:

$$p(r) = \begin{cases} \dfrac{r}{2a} + \dfrac{1}{2} & \text{if } -a \leq r \leq a \\ 1 & \text{if } a < r < 1-a \\ \dfrac{1-r}{2a} + \dfrac{1}{2} & \text{if } 1-a \leq r \leq 1+a \end{cases}$$

A plot of the probability density can be found in Figure S17. We note that $p(R)$ is continuous, since the probability densities agree with one another on the boundaries of the three regions, and of course $\int p(r)\mathrm{d}r = 1$. Interestingly, as $a \to 0$, $p(R)$ becomes a uniform distribution on $[0,1]$, and the resulting differential entropy, $h(R)$, limits to 0.

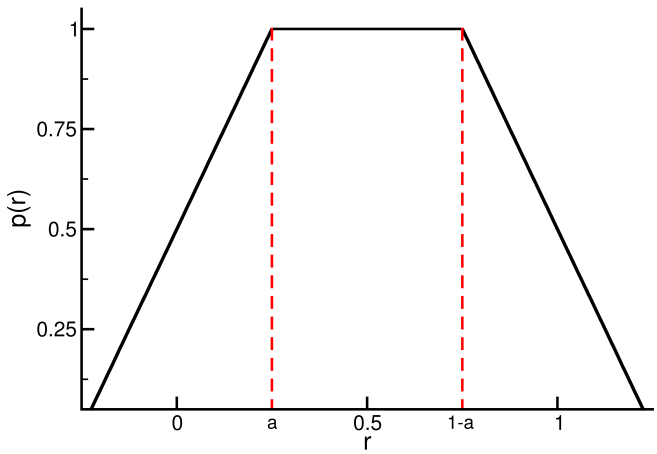

Figure S17: $p(r)$ when $a \leq < 12$. The red dotted lines delimit regions 1, 2, and 3 from left to right.

6.2.6 Calculating $h(R)$ and $C(S;R)$

With our expressions for $p_i(R)$ we can construct an expression for $h(R)$:

$$
\begin{aligned}
h(R) &= -\int_{-a}^{1+a} p(r) \log p(r) \, \mathrm{d}r \\
&= -\left[\int_{-a}^{a} p_1(r) \log p_1(r) \, \mathrm{d}r + \int_{a}^{1-a} p_2(r) \log p_2(r) \, \mathrm{d}r + \int_{1-a}^{1+a} p_3(r) \log p_3(r) \, \mathrm{d}r \right] \\
&= -\left[\int_{-a}^{a} \frac{r+a}{2a} \log \frac{r+a}{2a} \, \mathrm{d}r + \int_{a}^{1-a} 1 \log 1 \, \mathrm{d}r + \int_{1-a}^{1+a} \frac{1-r+a}{2a} \log \frac{1-r+a}{2a} \, \mathrm{d}r \right] \\
&= -\left[-\frac{a}{2} + 0 - \frac{a}{2} \right] \\
&= a.
\end{aligned}
$$

We also must have the conditional differential entropy, $h(R|S)$, to find the mutual information. Since $p(R|S)$ is a uniform distributions with a support range of $2a$, we know that:

$$\begin{aligned} h(R|S) &= -\int p(s) \cdot \int p(r|s) \log p(r|s)) \, \mathrm{d}r\mathrm{d}s \\ &= h(R|S=s) \cdot \int p(s) \, \mathrm{d}s \\ &= \log(2a) \cdot \int p(s) \, \mathrm{d}s \\ &= \log(2a) \cdot 1 \\ &= \log(2a). \end{aligned}$$

We then have an analytical expression for $C(S;R)$ when $a \leq \frac{1}{2}$:

$$I(S;R) = a - \log 2a.$$

6.3 Calculating $C(S;R)$ when $a \geq \frac{1}{2}$

We now calculate the expression for the channel capacity in the alternative scenario, when $a \geq \frac{1}{2}$.

6.3.1 Calculating $p(r)$

When $a \geq 1/2$, there is a region in R in which every value in S contributes to the probability density in that region of R. Correspondingly, there is a change in the signal bounds of the three regions of R values:

1. $-a \leq r \leq 1 - a$
2. $1 - a < r < a$
3. $a \leq r \leq 1 + a$.

These regions can be visualized in Figure S18.

6.3.2 Regions 1 & 3

One can easily show that the bounds of integration ($S_{m,i}(r)$ and $S_{M,i}(r)$) are the same in this scenario as they were for the scenario where $a < 0.5$. This results in identical expressions for the

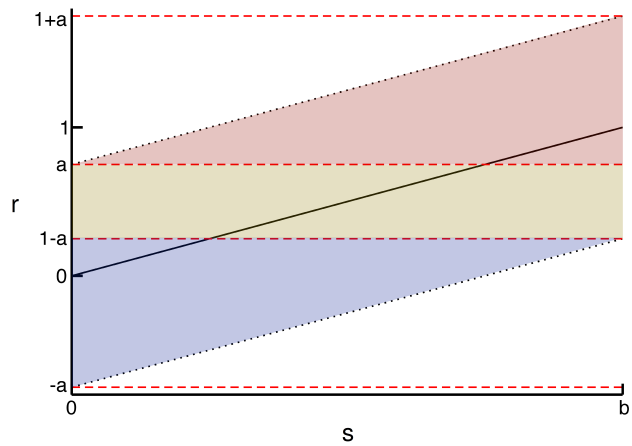

Figure S18: This plot shows a representation of a possible dose-response curve for the uniform model where $a \geq 1/2$. In this particular example, $a = 3/4$. The color scheme is as Figure S16. Note that the probability density in all of the response regions have contributions from all signal values, in contrast to the case where $a < 1/2$.

marginal probability densities:

$$p_1(r) = \frac{r}{2a} + \frac{1}{2}$$
$$p_3(r) = \frac{1-r}{2a} + \frac{1}{2}.$$

6.3.3 Region 2

This region is slightly different from that in 6.2.3 since all signals contribute equally to all response values (Figure S18). The bounds are thus the minimum and maximum signal values, 0 and b, respectively. The marginal probability is then:

$$p_2(r) = \int_0^b \frac{1}{2ab} \mathrm{d}s$$
$$= \frac{1}{2a}.$$

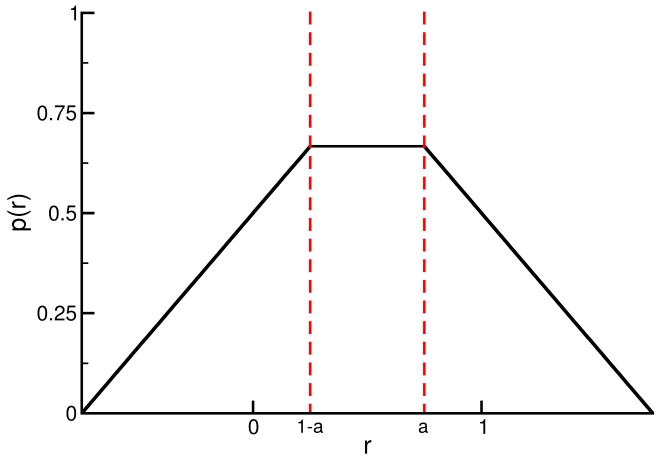

Figure S19: $p(r)$ when $a \geq 1/2$. The red dotted lines delimit regions 1, 2, and 3 from left to right.

6.3.4 Probability density of the response

As before we can assemble and plot the piecewise function for the probability density $p(r)$ (Figure S19):

$$p(r) = \begin{cases} \dfrac{r}{2a} + \dfrac{1}{2} & \text{if } -a \leq r \leq 1-a \\ \dfrac{1}{2a} & \text{if } 1-a < r < a \\ \dfrac{1-r}{2a} + \dfrac{1}{2} & \text{if } a \leq r \leq 1+a \end{cases}$$

Note that $p(r)$ is again continuous, and $\int p(r)\mathrm{d}r = 1$.

6.3.5 Calculating $h(R)$ and $C(S;R)$

As before we can now find the differential entropy of R and the channel capacity.

$$\begin{aligned} h(R) &= -\left[\int_{-a}^{1-a} \frac{r+a}{2a} \log \frac{r+a}{2a} \mathrm{d}r + \int_{1-a}^{a} \frac{1}{2a} \log \frac{1}{2a} \mathrm{d}r + \int_{a}^{1+a} \frac{1-r+a}{2a} \log \frac{1-r+a}{2a} \mathrm{d}r\right] \\ &= -\left[-\frac{1}{8a} - \frac{\log(2a)}{4a} - \left(\frac{1}{2a} - 1\right)\log(2a) - \frac{1}{8a} - \frac{\log(2a)}{4a}\right] \\ &= \frac{1}{4a} + \log(2a) \end{aligned}$$

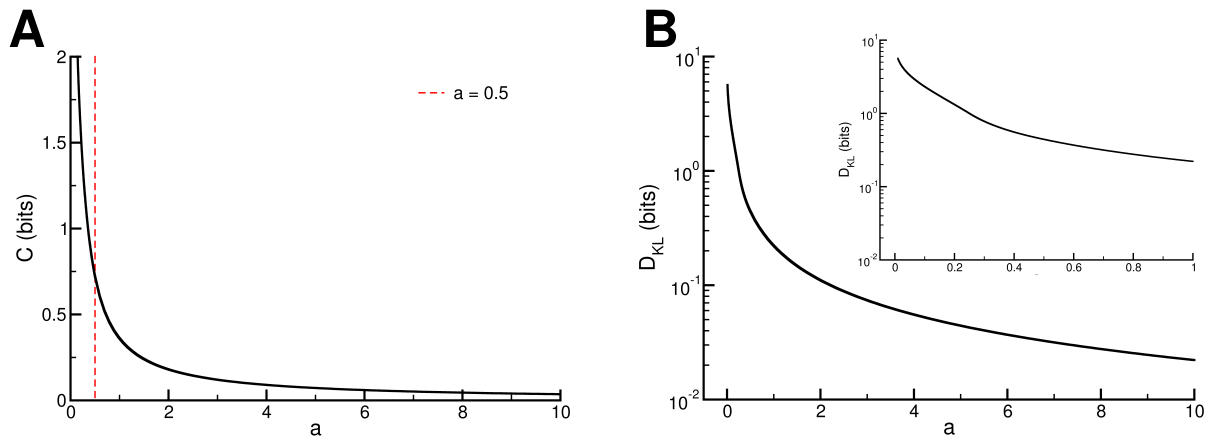

Figure S20: **A**) The channel capacity, $C(S;R)$, as a function of a. Note that the units of C have been converted to bits for the purposes of display. As expected, we observe that when $a \to \infty$, $C \to 0$. **B**) The Kullback-Leibler divergence between $p(r|s=K)$ and $p(r)$: $D_{\mathrm{KL}}(p(r|s=K)\,||\,p(r))$. Inset is the same plot but with a scale of 0 to 1 instead of 0 to 10.

Note that $h(R) \to \infty$ as $a \to \infty$, which is to be expected since the support of R becomes infinite in this case. Since the conditional differential entropy of R given S is identical to the model when $a < \frac{1}{2}$, we have mutual information for the model when $a \geq \frac{1}{2}$:

$$\begin{aligned} I(S;R) &= h(R) - h(R|S) \\ &= \frac{1}{4a} + \log(2a) - \log(2a) \\ &= \frac{1}{4a} \end{aligned}$$

6.4 Analysis of $C(S;R)$

From the above calculations we can construct a piecewise function which defines the channel capacity for this model (Figure S20A).

$$C(S;R) = \begin{cases} a - \log(2a) & \text{if } a < 0.5 \\ \dfrac{1}{4a} & \text{if } a \geq 0.5 \end{cases}$$

Note that when $a = \frac{1}{2}$, $a - \log(2a) = \frac{1}{4a}$, so $C(S;R)$ is continuous. Furthermore, the first and second derivatives of each portion of the piecewise function are equivalent when evaluated at $a = \frac{1}{2}$: the values are -1 and 4, respectively. The two components of the piecewise function do not, however, necessarily coincide for higher-order derivatives. Ultimately, however, we find that

the single-cell channel capacity decays as $1/a$ as a (or noise) increases above $1/2$ in this model.

As with the Gaussian version of the model, we can understand the above results by examining the Kullback-Leibler divergence between the conditional probability density of the response at half-maximal signal value (*i.e.* $p(r|s=K)$) and the marginal density $p(r)$. In this case, the simplicity of the model allows us to do this analytically:

$$\begin{aligned} D_{\mathrm{KL}}\left(p\left(r|s=K\right)||\,p\left(r\right)\right) &= \int p\left(r|s\right)\log\frac{p\left(r|s\right)}{p\left(r\right)}\mathrm{d}r \\ &= \int p\left(r|s\right)\log p\left(r|s\right)\mathrm{d}r - \int p\left(r|s\right)\log p\left(r\right)\mathrm{d}r \end{aligned}$$

where we have omitted the "$s=K$" terms from the conditional densities on the right-hand side of the equation for notational brevity. From here forward, we will use the term D_{KL} to denote this expression. Of course the first term is $-h\left(R|S=K\right) = -\log\left(2a\right)$, however the second term is slightly more complex. We know that $p\left(r|s=K\right)$ is only nonzero on the interval $[\frac{1}{2}-a, \frac{1}{2}+a]$, since $K=b/2$. This provides the bounds for the second integral above. These bounds, and the marginal probability density of r, depend on the value of a, and so to evaluate this integral, we again must split the model based on ranges of values of a. In this case there are three distinct intervals to consider:

1. $a \leq \frac{1}{4}$
2. $\frac{1}{4} < a \leq \frac{1}{2}$
3. $a > \frac{1}{2}$

In the first case, $p\left(r\right)=1$ and so the resulting expression for the Kullback-Leibler divergence becomes:

$$\begin{aligned} D_{\mathrm{KL}} &= -\log\left(2a\right) - \frac{1}{2a}\cdot\int_{\frac{1}{2}-a}^{\frac{1}{2}+a}\log 1\,\mathrm{d}r \\ &= -\log\left(2a\right) - \frac{1}{2a}\cdot 0 \\ &= -\log\left(2a\right) \end{aligned}$$

In the second case the integral must include terms to consider all regions of R space when

$\frac{1}{4} < a \leq \frac{1}{2}$:

$$D_{\mathrm{KL}} = -\log(2a) - \frac{1}{2a} \cdot \left[\int_{\frac{1}{2}-a}^{a} \log\left(\frac{r}{2a} + \frac{1}{2}\right) \mathrm{d}r + \int_{a}^{1-a} \log(1)\, \mathrm{d}r + \int_{1-a}^{\frac{1}{2}+a} \log\left(\frac{1-r}{2a} + \frac{1}{2}\right) \mathrm{d}r \right]$$
$$= -\log(2a) - \frac{1}{2a}\left[1 + 2\log 2 - \log\frac{1}{a} - 4a\right]$$

Finally, the third case includes all three regions of R when $a \geq \frac{1}{2}$:

$$D_{\mathrm{KL}} = -\log(2a) - \frac{1}{2a} \cdot \left[\int_{\frac{1}{2}-a}^{1-a} \log\left(\frac{r}{2a} + \frac{1}{2}\right) \mathrm{d}r + \int_{1-a}^{a} \log\left(\frac{1}{2a}\right) \mathrm{d}r + \int_{a}^{\frac{1}{2}+a} \log\left(\frac{1-r}{2a} + \frac{1}{2}\right) \mathrm{d}r \right]$$
$$= -\log(2a) - \frac{1}{2a} \cdot \left[-1 + \log\left(\frac{1}{a}\right) + (2a-1)\log\left(\frac{1}{2a}\right)\right]$$

Thus the Kullback-Leibler divergence over all values of a is:

$$D_{\mathrm{KL}} = \begin{cases} -\log(2a) & \text{if } a \leq \frac{1}{4} \\ -\log(2a) - \frac{1}{2a} \cdot \left[1 + 2\log 2 - \log\frac{1}{a} - 4a\right] & \text{if } \frac{1}{4} < a \leq \frac{1}{2} \\ -\log(2a) - \frac{1}{2a} \cdot \left[-1 + \log\left(\frac{1}{a}\right) + (2a-1)\log\left(\frac{1}{2a}\right)\right] & \text{if } a > \frac{1}{2} \end{cases}$$

and it can be visualized in Figure S20B. As with the Gaussian model, as $a \to \infty$, $D_{\mathrm{KL}} \to 0$, implying that $p(r)$ becomes equivalent to $p(r|s = K)$ for large a. We can see this intuitively in Figure S19: note that regions 1 and 3 have a constant "width" of 1, while the size of region 2 is $2a - 1$. As $a \to \infty$, region 2 grows without bound, whereas regions 1 and 3 do not and thus contribute less and less to the overall support for the density $p(r)$. In this limit, $p(r) \to \frac{1}{2a}$ on the interval $[1/2 - a, 1/2 + a]$, which is precisely $p(r|s = K)$.

7 Analytical model: Uniform signal and response distributions (cellular populations)

In this section, we use the simpler, uniform signal-response model to better understand how noise impacts information transfer to populations in the presence of a threshold.

7.1 Definitions

The following definitions are required for analysis of population-level information transmission:

- θ is the decision-making threshold; cells with $r \geq \theta$ die and those with $r < \theta$ survive.
- N is total the number of cells in a population
- $p_d(s)$ is the probability of some cell in the population dying given a particular signal value
- N_D is the total number of dead cells (*i.e.* the population response)

Other variables, such width of the uniform probability distribution (a) are identical to the single-cell model (Section 6). For the following analysis we assume that $\theta = \frac{1}{2}$ to simplify the analysis. As in our single-cell analysis, there are two cases that produce distinct behaviors:

1. $0 < a \leq \dfrac{1}{2}$
2. $a > \dfrac{1}{2}$

7.2 Characterizing signal space

7.2.1 Case 1

Since $a \leq \frac{1}{2}$, there is always a region in signal space near $s = 0$ in which all cells fall below the threshold (they all survive) that can be seen in Figure S21. The upper bound of this region is termed $S_{\min}$. Similarly, there is a region near $s = b$ where all cells are above the threshold (they all die), whose lower bound is $S_{\max}$. These values can also be defined in such a way that $S_{\min}$ is the largest value of s for which no cells die, and $S_{\max}$ is the smallest value of s for which all cells die. These two bounds delimit the region of signal space that we term the *transition zone*, and can be determined algebraically. We previously characterized the single-cell response, r, given some value $s \in S$ as a uniform distribution that lies on the interval:

$$r(s) \in \left[\frac{s}{b} - a, \frac{s}{b} + a\right]$$

Thus $S_{\min}$ corresponds to the signal producing the response distribution whose maximum is equal to the decision threshold:

$$\frac{1}{2} = \frac{S_{\min}}{b} + a$$

We can find $S_{\max}$ similarly, resulting in a transition zone corresponding to the interval:

$$[S_{\min}, S_{\max}] = \left[b\left(\frac{1}{2} - a\right), b\left(\frac{1}{2} + a\right)\right]$$

which can be visualized in Figure S21. This region of signal space is of interest since it is the only region that positively contributes to the population mutual information by having a nonzero correlation between S and the number of dead cells in a population (the population response, N_D) as seen in Figure S22A. We will restrict our analysis of the population information transmission in case 1 to this particular region of signal space.

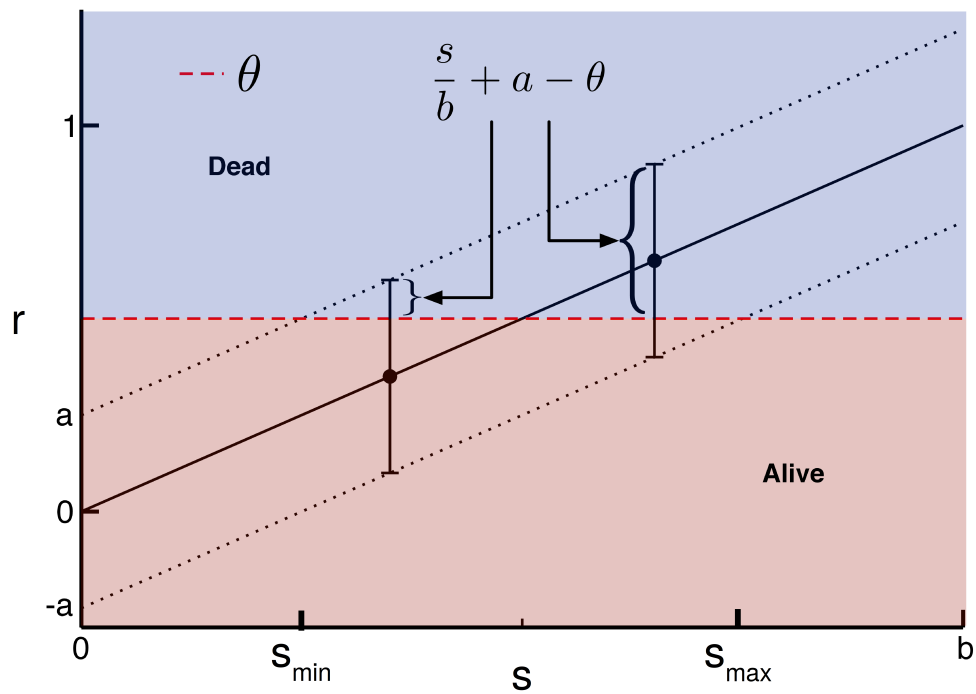

Figure S21: A visual representation of population response to signal in terms of fractional survival of cells, where θ is the threshold above which cells die; the other variables are defined in the text. We can find the fraction of cells that die given a particular signal value with a simple algebraic expression $s/b + a - \theta$ and then normalize it to define the probability of death for a cell exposed to a signal value, defined in the text as: $p_d(s)$.

7.2.2 Case 2

Since $a > \frac{1}{2}$, all signal values produce response distributions that generate fractional population survival. Therefore the bounds $S_{\min}$ and $S_{\max}$ do not exist, since they would be less than 0 and greater than b, respectively. In other words, the entirety of signal space considered falls within the transition zone.

7.3 Population response to signal

Given the definitions and limits defined above, we can construct both the conditional response of the population (i.e. $p(N_D|s)$) and, from that distribution, calculate the average response as a function of signal. Given the conditional response of individual cells to signal (*i.e.* $p(r|s)$), we can calculate the probability that a cell in the population dies, based on the proportion of the response distribution that lies above the decision-making threshold, $\theta = \frac{1}{2}$, as seen in Figure S21:

$$p_d(s) = \int_{\theta}^{\frac{s}{b}+a} p(r|s)dr = \frac{\frac{s}{b} + a - \frac{1}{2}}{2a}$$

where we have implicitly assumed that $S_{min} \leq s \leq S_{max}$ for the cases defined above, depending on the value of a. In our model, all cells respond independently, so it follows that the number of cells that die given a particular population size and signal value is binomially distributed:

$$p(N_D|s) = \binom{N}{N_D} p_d(s)^{N_D} (1 - p_d(s))^{N-N_D}$$

As a result, the mean number of dead cells at any signal value s is:

$$\overline{N_D}(s) = p_d(s) \cdot N$$

which defines the (average) signal-response behavior of the population.

7.3.1 Average population response for Case 1

As discussed above, we have either complete survival and or complete death in populations exposed to signals below $S_{\min}$ and above $S_{\max}$, respectively (see Figure S22A). Given this fact, we obtain the following piecewise function for $\overline{N_D}(s)$:

$$\overline{N_D}(s) = \begin{cases} 0 & \text{if } s < S_{\min} \\ \dfrac{\frac{s}{b} + a - \frac{1}{2}}{2a} \cdot N & \text{if } S_{\min} \leq s \leq S_{\max} \\ N & \text{if } s > S_{\max} \end{cases}$$

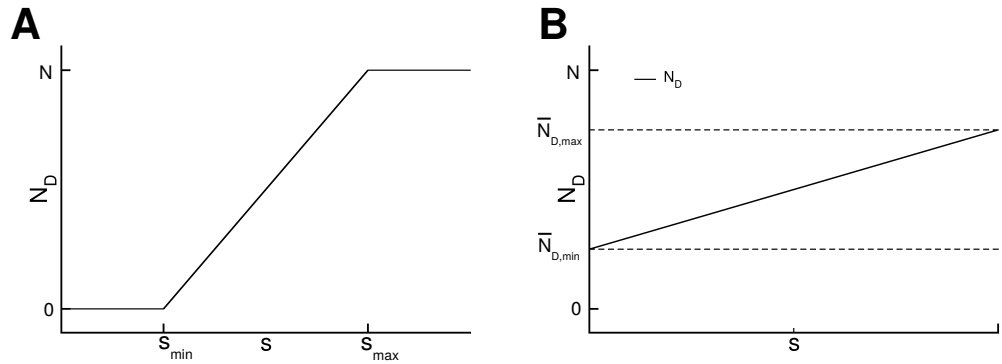

Figure S22: A plot of mean population response to signal for case 1 (**A**) and case 2 (**B**)

7.3.2 Case 2

The bounds of signal space in this case correspond to instances in which the population response is fractional (Figure S22B). We can define the minimum and maximum average population response as follows:

$$\overline{N_D}_{,\min} = N\left(\frac{a-\frac{1}{2}}{2a}\right)$$
$$\overline{N_D}_{,\max} = N\left(\frac{1+a-\frac{1}{2}}{2a}\right) = N\left(\frac{a+\frac{1}{2}}{2a}\right)$$

In this case, the average population response follows:

$$\overline{N_D}\left(s\right) = \frac{\frac{s}{b}+a-\frac{1}{2}}{2a}\cdot N = \overline{N_D}_{,\min} + \frac{N}{2ab}s$$

7.4 Finding the optimal signal distributions

As before, to calculate the maximum of I (*i.e.* the channel capacity), we define the probability density of S based on the first derivative of the average signal-response function. To do this, however, we must focus on a dose-response relationship that is defined on the interval $(0, 1)$, so we focus on the *fractional* response of the population. The form of this fractional response depends on the value of a as described below.

7.4.1 Case 1

Taking this derivative is relatively simple for case 1, since the minimum and maximum fractional responses are 0 and 1, respectively, and so we can define the fractional response as just $\frac{\overline{N_D}}{N}$ in this

case. Thus, for $s \in [S_{min}, S_{max}]$, we have:

$$p\left(s\right) \equiv \frac{\mathrm{d}\frac{\overline{N_D}}{N}}{ds} = \frac{1}{2ab}$$

Note that this is just the uniform probability density on $[S_{min}, S_{max}$. Outside the transition zone, $\overline{N_D}$ does not depend on s, so $p(s) = 0$.

7.4.2 Case 2

As mentioned above, interpreting the average dose-response function as the CDF of the probability density of signal requires that this average cover the interval $(0, 1)$. Finding $p(s)$ thus requires care when $a > 1/2$ since the minimum and maximum population responses are not 0 and 1, and depend on a. To overcome this issue, we defined a normalized response function ($R'\left(s\right)$) that has minimum and maximum values of 0 and 1, respectively:

$$R'\left(s\right) \equiv \frac{\overline{N_D}\left(s\right) - \overline{N_D}_{,\mathrm{min}}}{\overline{N_D}_{,\mathrm{max}} - \overline{N_D}_{,\mathrm{min}}}$$

which simplifies to

$$R'\left(s\right) = \frac{s}{b}$$

The resulting signal density is then just:

$$p\left(s\right) \equiv \frac{\mathrm{d}R'}{ds} = \frac{1}{b}$$

which again is just the uniform density on the relevant domain of S (*i.e.* $[0, b]$).

7.5 Calculating population-level channel capacities

With the above signal distributions, we can calculate the channel capacity between the signal and response (as the number of dead cells) by determining the marginal and conditional entropies:

$$C\left(S; N_D\right) = H\left(N_D\right) - H\left(N_D|S\right)$$

7.5.1 Case 1

Finding $H\left(N_D|S\right)$ is relatively straightforward, since the approximate entropy of a binomial distribution is known:

$$\begin{aligned}
H\left(N_D|S\right) &= \int_{S_{\min}}^{S_{\max}} p\left(s\right) \cdot \left(\sum_{N_D=0}^{N} -p\left(N_D|S\right) \log p\left(N_D|S\right) \right) \mathrm{d}s \\
&\approx \int_{S_{\min}}^{S_{\max}} \frac{1}{2ab} \cdot \frac{1}{2} \log\left(2\pi \mathrm{e} \cdot N \cdot p_d\left(s\right)\left(1 - p\left(s\right)\right)\right) \mathrm{d}s \\
&= \frac{1}{2} \log\left(2\pi N\right) - \frac{1}{2}
\end{aligned}$$

We then need the marginal probability of the number of dead cells to calculate $H\left(N_D\right)$:

$$\begin{aligned}
p\left(N_D\right) &= \int_{S_{\min}}^{S_{\max}} p\left(s\right) \cdot p\left(N_D|s\right) \mathrm{d}s \\
&= \int_{S_{\min}}^{S_{\max}} \frac{1}{2ab} \cdot \binom{N}{N_D} p_d\left(s\right)^{N_D} \left(1 - p_d\left(s\right)\right)^{N-N_D} \mathrm{d}s \\
&= \frac{1}{1+N}
\end{aligned}$$

Interestingly, for case 1, we thus have that the marginal probability $p(N_D)$ is just the uniform probability mass function for the set $\{0, 1, \ldots, N\}$; in other words, the total probability of finding N_D dead cells across all of the relevant signal space is just the uniform probability over all the possible states (ranging from 0 cells dying to all of them dying). It is then straightforward to calculate the entropy:

$$H\left(N_D\right) = \sum_{N_D=0}^{N} -\frac{1}{1+N} \log \frac{1}{1+N} = \log\left(1+N\right)$$

Thus when $0 < a \leq \frac{1}{2}$ we find that

$$C\left(S; N_D\right) = \log\left(1+N\right) - \left(\frac{1}{2} \log\left(2\pi N\right) - \frac{1}{2} \right) \tag{4}$$

We first note that this expression is invariant on a for $a \leq \frac{1}{2}$. This echoes our findings in Section 4.5: since we have taken S to be continuous, then by definition the "minimal signal resolution" $\Delta s = 0$. As a consequence, the population-level channel capacity does not depend on the level of noise below the $a = \frac{1}{2}$ threshold. Also, note that, as N becomes large, we can re-write this

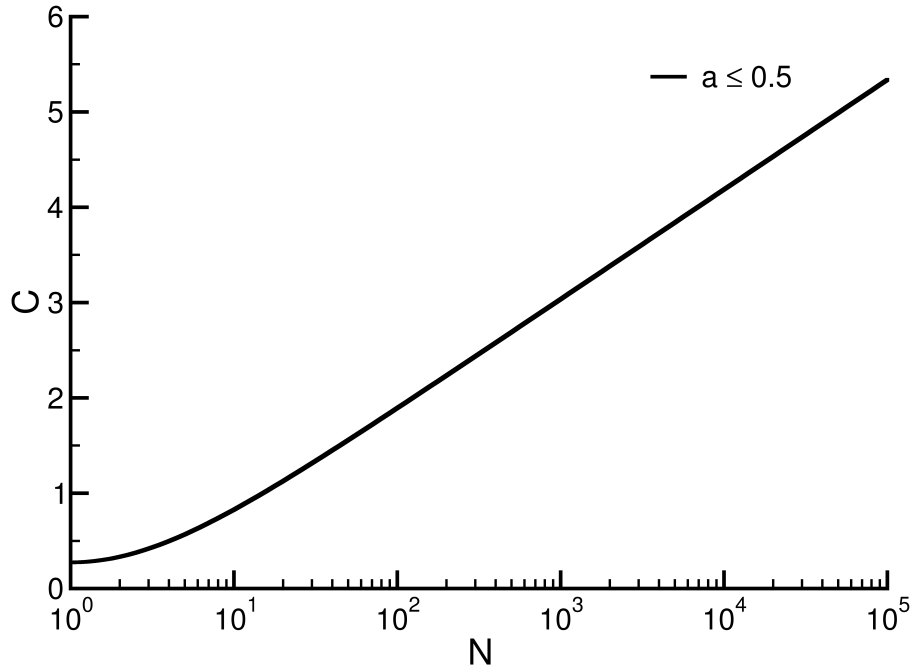

Figure S23: $C\left(S; N_D\right)$ increases monotonically with the size of the population, N, when $a \leq 1/2$ (Equation 4). For large values of N this increase follows $C(S; N_D) \sim \log \sqrt{N}$ (Equation 5).

expression as:

$$C(S; N_D) \sim \frac{1}{2} \log N \tag{5}$$

or, in other words, the channel capacity is just $\log \sqrt{N}$. This is perhaps not surprising, given the underlying binomial structure of $p(N_D|s)$, but nonetheless indicates that the maximum of the population-level information transfer scales monotonically with N (Figure S23 and Figure 3D in the main text).

7.5.2 Case 2

The expression for conditional entropy is slightly more complicated when $a > \frac{1}{2}$:

$$\begin{aligned} H\left(N_D|S\right) &= \int_0^b \frac{1}{b} \cdot \left(\frac{1}{2}\log\left(2\pi \mathrm{e} \cdot N \cdot p_d\left(s\right)\left(1 - p_d\left(s\right)\right)\right)\right) \mathrm{d}s \\ &= \frac{1}{2}\log\left(2a+1\right) - \frac{1}{2} + a\log\left(2a+1\right) + \frac{1}{2}\log\left(2a-1\right) - a\log\left(2a-1\right) \\ &\quad - \frac{3}{2}\log\left(2\right) + \frac{1}{2}\log\left(\pi\right) - \log\left(a\right) + \frac{1}{2}\log\left(N\right) \end{aligned}$$

Again, finding the marginal entropy requires an expression for $p\left(N_D\right)$:

$$\begin{aligned} p\left(N_D\right) &= \int_0^b \frac{1}{b} \cdot \binom{N}{N_D} p_d\left(s\right)^{N_D}\left(1 - p_d\left(s\right)\right)^{N-N_D} \mathrm{d}s \\ &= 2a \cdot \binom{N}{N_D} \cdot \left(\mathrm{B}_u\left(N_D+1, N-N_D+1\right) - \mathrm{B}_v\left(N_D+1, N-N_D+1\right)\right) \end{aligned}$$

where $\mathrm{B}_x\left(a,b\right)$ denotes the incomplete beta function and

$$\begin{aligned} u &= \frac{a+\frac{1}{2}}{2a} \\ v &= \frac{a-\frac{1}{2}}{2a} \end{aligned}$$

Interestingly, this expression mimics the one we obtained for single-cell responses in the Gaussian model (Section 5). Calculating $H(N_D)$ requires computing the following finite sum:

$$H(N_D) = -\sum_{N_D=0}^{N} p(N_D)\log p(N_D) \tag{6}$$

Computing this sum in closed form is made difficult by the fact that the difference of incomplete beta functions enters the expression under the log, and in the end we did not discover a solution. We therefore numerically calculated the marginal entropy, $H\left(N_D\right)$ for values of $a \in \left[\frac{1}{2}, 2\right]$, and plotted our results for the mutual information together with those from case 1 for multiple population sizes (Figure S24). We see that, for $a > \dfrac{1}{2}$, the population-level information decreases significantly as a increases (see Figure S24). Intuitively, this occurs because the transition in N_D no longer covers the entire set of possibilities from 0 to N (Figure S22B).

8 Experimental Methods

HeLa cells were maintained in DMEM medium (Corning 10-013-CV) with 10% fetal bovine serum and 1% penicillin/streptomycin solution (Life Technologies 15140-122). For TRAIL dose-response assays, HeLa cells were plated at a density of 250k cells/well in 12-well plates (Sigma SIAL0513), allowed to adhere overnight, and treated with varying doses of SuperKiller TRAIL (Axxora

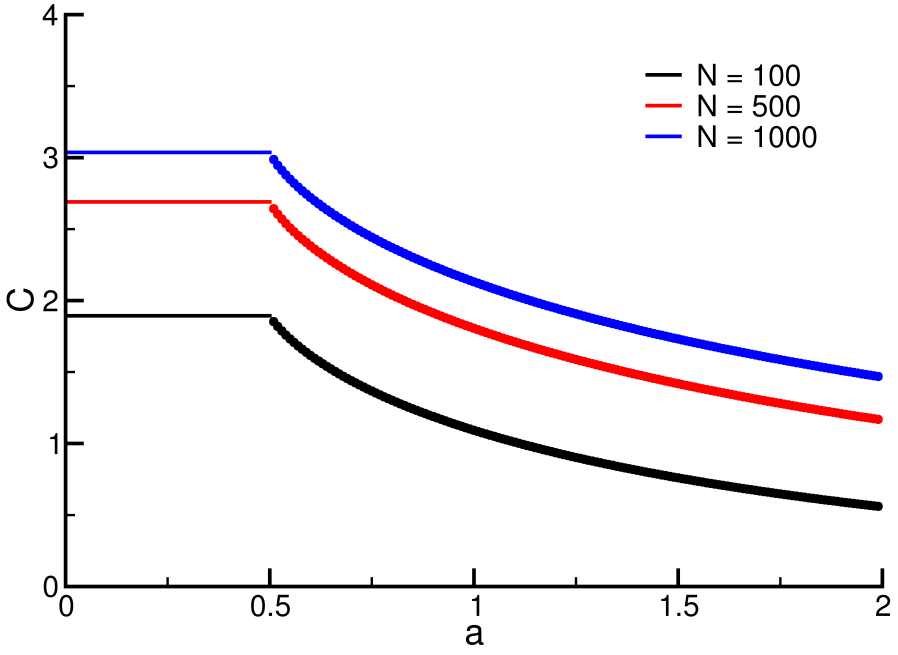

Figure S24: $C\left(S;N_D\right)$ as a function of a when $\theta = 1/2$. For $a < 1/2$ we used the closed-form solution from our model (Equation 4). For $a \geq 1/2$ we numerically calculated C by directly evaluating the sum in Equation 6. Note that C is invariant on a for $a < 1/2$, but then decreases monotonically as a increases above that threshold.

ALX-201-115) for 11 hours. Three replicate wells were used for each TRAIL dose to establish the technical variability of the assay. After treatment, medium containing dead cells was transferred to flow cytometry tubes (BD Falcon 352235) containing 2 ml FACS buffer (PBS + 10% fetal bovine serum); cells remaining in the wells were removed by trypsinization, added to the corresponding tubes, pelleted by centrifugation and fixed in 4% paraformaldehyde for 30 minutes. After fixation cells were washed twice in PBS and permeabilized in 100% methanol overnight at -20C. Cells were stained with primary antibodies to cleaved caspase 3 (rabbit anti-cleaved caspase 3, BD 559565) and cleaved PARP (mouse anti-cleaved PARP, BD 552596) 1:250 in FACS buffer (PBS + 0.1% Tween-20) for 1 hour at 25C. Cells were washed twice in PBS-T, then treated with secondary antibodies: Alexa-488 donkey anti-rabbit IgG (Life Technologies A-21206) and Alexa-594 donkey anti-mouse IgG (Life Technologies A21203), 1:500 in FACS buffer for 1 hr at 25C. Cells were washed in PBS-T, resuspended in PBS, and counted on a flow cytometer (BD LSRII), with 20,000 cells analyzed per experimental replicate.

MCF10A cells were obtained from J. Brugge (Harvard Medical School, Boston, MA) and cultured as described (22). For TRAIL dose response assays, MCF10A cells were plated in 96-well plates (Corning 353072) and treated with varying doses of SuperKiller TRAIL for 11 hours. After treatment the cells were washed with PBS and the density of viable cells was assayed by methylene blue staining as described previously (23).

References

1. Shannon CE (1948) A Mathematical Theory of Communication. *Bell System Technical Journal* 27:379–423.

2. Paninski L (2003) Estimation of Entropy and Mutual Information. *Neural Computation* 15(6):1191–1253.

3. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A (2011) Information transduction capacity of noisy biochemical signaling networks. *Science (New York, NY)* 334(6054):354–358.

4. Strong S, Koberle R, de Ruyter van Steveninck R, Bialek W (1998) Entropy and Information in Neural Spike Trains. *Physical Review Letters* 80(1):197–200.

5. Steuer R, Kurths J, Daub CO, Weise J, Selbig J (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics (Oxford, England)* 18 Suppl 2:S231–40.

6. Albeck JG, Burke JM, Spencer SL, Lauffenburger DA, Sorger PK (2008) Modeling a Snap-Action, Variable-Delay Switch Controlling Extrinsic Cell Death. *PLoS Biology* 6(12):e299.

7. Albeck JG et al. (2008) Quantitative Analysis of Pathways Controlling Extrinsic Apoptosis in Single Cells. *Molecular cell* 30(1):11–25.

8. (2010) *R: A Language and Environment for Statistical Computing.* (The R Foundation for Statistical Computing, Vienna, Austria).

9. Bashor CJ, Helman NC, Yan S, Lim WA (2008) Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics. *Science (New York, NY)* 319(5869):1539–1543.

10. Sacan A, Ferhatosmanoglu H, Coskun H (2008) CellTrack: an open-source software for cell tracking and motility analysis. *Bioinformatics (Oxford, England)* 24(14):1647–1649.

11. Burov S et al. (2013) Distribution of directional change as a signature of complex dynamics. *PNAS* 110(49):19689–19694.

12. van der Wath RC, Gardiner BS, Burgess AW, Smith DW (2013) Cell organisation in the colonic crypt: A theoretical comparison of the pedigree and niche concepts. *PLoS ONE* 8(9):1–15.

13. Lencer W et al. (1997) Induction of epithelial chloride secretion by channel-forming cryptdins 2 and 3. *PNAS* 96(16):8585–9.

14. Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427:415–8.

15. Hilfinger A, Paulsson J (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *PNAS* 108(29):12167–72.

16. Elowitz MB (2002) Stochastic Gene Expression in a Single Cell. *Science (New York, NY)* 297(5584):1183–1186.

17. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS* 99:12795–12800.

18. Huber MF, Bailey T, Durrant-Whyte H, Hanebeck UD (2008) On entropy approximation for gaussian mixture random vectors in *Proc IEEE Int Conf Multisensor Fusion Integr Intell Syst*. pp. 181–188.

19. Laughlin S (1981) A simple coding procedure enhances a neuron's information capacity. *Zeitschrift fur Naturforschung - Section C Journal of Biosciences* 36(9-10):910–912.

20. Wolfram Research, Inc. (2016) Mathematica.

21. Oliphant TE (2007) Python for scientific computing. *Computing in Science & Engineering* 9:10–20.

22. Debnath J, Muthuswamy SK, Brugge JS (2003) Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* 30(3):256–268.

23. Flusberg DA, Roux J, Spencer SL, Sorger PK (2013) Cells surviving fractional killing by TRAIL exhibit transient but sustainable resistance and inflammatory phenotypes. *Molecular Biology of the Cell* 24(14):2186–2200.