Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Cellular noise and information transmission

Andre Levchenko[1] and Ilya Nemenman[2,3]

The technological revolution in biological research, and in particular the use of molecular fluorescent labels, has allowed investigation of heterogeneity of cellular responses to stimuli on the single cell level. Computational, theoretical, and synthetic biology advances have allowed predicting and manipulating this heterogeneity with an exquisite precision previously reserved only for physical sciences. Functionally, this cell-to-cell variability can compromise cellular responses to environmental signals, and it can also enlarge the repertoire of possible cellular responses and hence increase the adaptive nature of cellular behaviors. And yet quantification of the functional importance of this response heterogeneity remained elusive. Recently the mathematical language of information theory has been proposed to address this problem. This opinion reviews the recent advances and discusses the broader implications of using information-theoretic tools to characterize heterogeneity of cellular behaviors.

Addresses
[1] Yale Systems Biology Institute and Department of Biomedical Engineering, Yale University, New Haven, CT 06520, USA
[2] Department of Physics, Emory University, Atlanta, GA 30322, USA
[3] Department of Biology, Emory University, Atlanta, GA30322,USA

Corresponding author: Levchenko, Andre (andre.levchenko@yale.edu)

Current Opinion in Biotechnology 2014, **28**:156–164

This review comes from a themed issue on **Systems biology**

Edited by **Christian M Metallo** and **Victor Sourjik**

http://dx.doi.org/10.1016/j.copbio.2014.05.002

Biological systems, including cells and tissues, are extremely diverse in nature. This diversity can frequently be perplexing, particularly if displayed by cells of the same type and carrying identical genomes. Indeed, if the genetic composition of a cell of a certain type fully defines its phenotypic responses to a specific environmental input, then cells of the same type would be expected to show similar responses to identical stimuli. However, this is rarely the case. For most cells, there is an easily observable variability in their responses, even if the cells are clonal in origin. This variability is sometimes referred to as biological noise. The sources, the analysis, and biological, medical, and biotechnology implications of this noise are the focus of this perspective. We posit that noise should not only be acknowledged, but measured and predicted with quantitative accuracy, and that the mathematical language of information theory is the right framework for characterization of noisy responses of cellular systems.

Emergence of biological noise in single cell behavior

The recent technological revolution in biological research brought with it a diverse array of methods allowing one to have a high-resolution view of single cell behaviors. In addition to the analysis of cellular phenotypes, such as cell division, migration or death, these methods allow detection of the biochemical events underlying cellular decision-making leading to various defined phenotypic outcomes [1]. A particularly important technique has been the use of molecular fluorescent labels [2]. Such labels allow tracking individual molecules in both live and fixed cells, revealing time dependent molecular activation and localization patterns. Over time, use of these tools revealed that both a given cell behavior and the underlying biochemical processes are highly variable in a way not immediately interpretable [3–6]. This variability can often be traced to an observation that chemical reactions in cells occur with very low numbers of molecules (one copy of DNA, tens of copies of regulatory proteins in bacteria [7], and so on [8]). Such reactions result in unpredictably fluctuating numbers of molecules in individual cells or their compartments, and thus in different effective chemical concentrations across cellular populations. This variability is similar to the famed shot noise in electronic devices [9]. Thus it is not surprising that understanding this cellular variability has required new theories and approaches that have their roots in physical sciences. These new ideas have come in two favors: mathematical and computational tools for efficient treatment of noise in biochemical processes, and a series of novel experimental techniques focused on measuring the noise.

The *computational advances* in understanding cellular noise started with the Stochastic Simulation Algorithm, also known under the name of its developer, Dan Gillespie [10,11]. This algorithm captures the stochasticity of biochemical reactions, generating statistically accurate random time courses of concentrations of reacting chemicals. Its simplicity was in stark contrast to generally intractable analytical approaches that had been used traditionally [9]. Various extensions of the algorithm followed, making it computationally more efficient under different conditions (see e.g., [11–14]), and applicable to very large biochemical reaction networks [15]. At the

same time, mathematical advances have resulted in powerful analytical techniques for modeling noise [16–18]. Similarly, the typically linear nature of the dynamics of probabilities of molecule numbers gave rise to efficient numerical approaches [19].

All of these methods are now mature enough to make accurate predictions within their domains of applicability. However, their assumptions often fail in the context of real living cells. For example, the assumption of well-mixed chemical reactions may not hold for physically structured cells [20], and the assumption of reactions happening as independent, almost instantaneous events may fail for enzymes with complex kinetics [21]. Further, such stochastic modeling is data hungry — generally, predictions are sensitive to values of many kinetic parameters, which must be inferred from data. Thus some recent computational developments have focused on identifying 'essential' features of stochastic dynamics and building their coarse-grained yet accurate models [22–24].

The *experimental analysis* of the noise in cellular responses has focused primarily on diversity in expression of individual genes, and often, in particular, on the relative contributions of so-called *intrinsic* and *extrinsic* noise components [25]. More specifically, careful introduction of two different fluorescent labels to track the transcription of a gene of choice decomposes the expression variability into components co-variable or independent between the two labels [4,26]. The co-varying component reports on the cell–cell variability that has common effects on expression of many genes. The independent variability of the labels corresponds to the intrinsic stochasticity in expression of specific genes of interest. Although one component may dominate the other under different circumstances, or for distinct genes or cell types, a very important result of this analysis is that the noise can be complex and multifactorial in its origins.

More complex experimental designs have suggested that noise can accumulate within gene regulation cascades, in which products of gene expression reactions are themselves regulators that can control reactions [27,28]. Products in such chains may feed back and modulate the activity of the earlier biochemical species, forming networks that endow cellular noise with complex and yet not fully understood statistical properties. Here gene expression noise may produce qualitatively novel behaviors, such as switching among multiple phenotypic states or transient activation [7,29–32]. Importantly, the developed computational methods can predict the mean responses and the fluctuations for these relatively complex networks [33,34,35•]. Some examples of noise analysis in live cells are described in excellent recent reviews, including [1,6,36,37].

In summary, intrinsic variability of cell responses is a pervasive characteristic of single cell behavior. It now can be measured, modeled, and sometimes manipulated with an exquisite precision previously reserved only for physical sciences. And yet it is not obvious how to quantify the functional importance of the observed cell-to-cell variability. Does the variability compromise certain responses by a cell to signals from the environment, and hence is detrimental to the cell's function? Or might this variability be embraced by the cells because it enlarges the repertoire of possible responses and hence assist a group of cells in increasing the adaptive nature of their behavior [36]? These questions are particularly important in the context of the interests encompassed by modern biotechnology. Indeed, can diversity of cell responses to the specific drug compromise its effects? Given the cellular noise, how robust would the performance of new synthetically engineered organisms be? We discuss these considerations next.

The dose response and implications of biological noise

Diversity of cellular states and thus cellular responses to extrinsic signals suggests that the overwhelmingly deterministic view of biological systems needs to undergo a dramatic change. This paradigm shift is not unlike the shift from the determinism of classical physics to the essential uncertainty of quantum mechanics. Inevitably, such tectonic shifts are fraught with initial doubt and hesitance, but can also bring about a completely new set of research questions, while shedding light on many important unresolved problems. The ostensibly simple and common dose response assay can be the case in point.

The common interpretation of a typical dose response of a biochemical reaction is that a graded increase in abundance of an active compound (enzyme, substrate, inhibitor, etc.) leads to a graded change in the output of the reaction. This view persists from college textbooks, and it is reinforced daily by typical biochemical assays, including those performed in biotechnology and pharmaceutical companies. Such essays suggest saturating, possibly non-linear, sometimes bimodal, but mostly very smooth input–output relationships. These imply that a small change in the input can be sensed and converted into a small change in the output over wide ranges of both. Thus many (perhaps infinitely many) distinct doses can be accurately converted into equally many distinct response levels. Conversely, by observing a response, one of the many distinct doses can be inferred. An alternative picture is also common, especially when analyzing emergence of distinct cellular phenotypes. Here the dose–response curve is discontinuous, and sometimes even hysteretic [7,38,39•]. However, a change of the dose across a certain threshold turns a gene 'on' or 'off' in a precise, deterministic fashion reminiscent of digital computers. In both pictures, the level of precision paints

chemical reactions as versatile and predictable devices that can be controlled, for example, by various drugs with a considerable certainty.

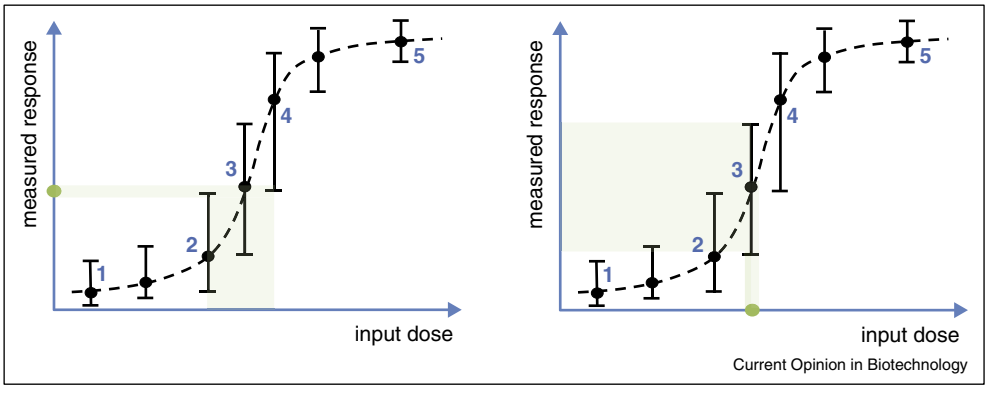
Such high fidelity view of biochemistry not only in the test tube, but also in live cells, can increase one's confidence in the effects of drugs within a cell population and, ultimately, in an organism. However, frequently this view is misleading. Indeed, most of the 'classical' biochemical assays rely on large amounts of material, and thus thousands or millions of individual cells. Their results represent the average view of a cell population response, hiding the details of the individual cell behavior. Can the existence of this individuality, the biological noise, change this interpretation of the dose response curve for biochemical processes?

The average, population-level response to a change in the cellular environment is an incomplete description of a population behavior. In the worst case, when individual cells respond in a bimodal fashion (i.e., have clear 'on' and 'off' response states), the mean population response may not be representative of *any* cell in the population [7] (similar caution has been sounded in other branches of biology, see e.g., [40]). However, even in the best case when the mean population response is representative of a typical individual response, a more statistically informative view depends on additional population characteristics, such as the response variance. Indeed, population responses with different degrees of variance can still have the same average values. However, a higher variance implies the existence of responses that can be substantially higher or lower than the average. Thus some cells might show a very high response, and others virtually no response at all when faced with the same concentration of a drug or a growth factor. Overall, predictability of a response to a signal of any given cell within a population decreases as the population variability goes up, see Figure 1 for an illustration.

The distinction between such low and high noise situations may be extremely biological significant. For instance, if a drug is designed to target cells that can quickly proliferate, a high diversity of cell responses will leave some relatively unresponsive cells unscathed. This will select for these 'drug-resistant' cells and thus may decrease the sensitivity to the follow-up treatments. Antibiotic persistence [38,39,41–44] and cancer stem cells [45] are examples of such phenomena.

Limitations on predictability of responses of individual cells due to biological noise demand reassessment of the idea of potentially infinite sensitivity of a generic dose response. Indeed, the inability to predict response of a single cell means that it is also impossible to infer which dose was used to stimulate a given cell from simply observing its response. This raises the important question of whether the *response* noise can effectively 'drown out' the *input* signal in the biological signaling. To illustrate this idea one can resort to the familiar concept of error bars, see again Figure 1. The rule of thumb is that if two data points have substantially overlapping error bars (designating, for example, the 95% confidence intervals), then the difference between these two data points is not statistically significant. In the worst case scenario, all data points within the dose response can have overlapping error bars, making any dose dependence suggested by the mean behavior statistically insignificant. Suppose, however, that the extreme levels of response (the lowest and the highest response values)

Figure 1



(Left) Mean population responses are typically captured by the dose–response curve. However, because of the population variability, denoted here by error bars, the dose response provides a lot less information than would be expected from a deterministic relation. In particular, responses to doses 2 and 3 or 3 and 4 overlap. Thus, for example, observing the response near the mean of the dose 3 does not allow to identify reliably which dose has caused it, and all doses between 2 and 4 are possible (green shaded areas). On the contrary, responses to doses 1 and 4 or 2 and 5 do not overlap, and these can be distinguished. Information theory allows one to quantify an effective number of such distinguishable dose–response pairs in this and similar systems. (Right) Inability to infer the dose that has led to the response is equivalent to inability to predict the response from the dose. Here the dose 3 leads to responses that span the range of mean responses from dose 2 to almost 4 (green shaded area).

are statistically distinguishable. This implies that at least two doses can be distinguished no matter which cells in the lowest and highest response ranges are selected. But can three or more doses be distinguished as well? This question cannot be easily addressed for cases with high levels of noise, that is, for data with large enough error bars. Clearly, one no longer can expect accurate sensing of a great number of different doses, as would be suggested by average responses. However, it is also not immediately clear whether the number of distinguishable doses is, say 2, 4, or perhaps 3.5. To answer this question, one needs a specifically designed mathematical approach. Fortunately, the appropriate language and the corresponding analysis methods exist and have been used with a great success in a variety of fields, from communication theory, to physics, and to neuroscience. This language is *information theory*.

Quantifying biological noise with information theory

Information theory has been developed originally by Shannon to quantify communication through a noisy communication channel [46]. One of its fundamental results is that reliable communication over a noisy channel is possible if the amount of data to be transmitted is below a certain channel-dependent threshold [47,48]. More recent developments have extended the formalism to communications among a complex network of agents [49], but large parts of this theory are still a work in progress [50,51]. The questions addressed by information theory are similar to those raised above in our analysis of the dose response: what, or how much, can be said reliably about the signal(s) when observing a response(s) to them? An attractive aspect of the mathematical apparatus underlying information theory is that it is applicable regardless of the nature of the signal, of the medium transmitting it, or of the communication noise. This makes the language of Shannon's information theory immediately useable for a quantitative analysis of biological signal transduction.

Shannon's theory is intrinsically probabilistic, and the amount of communicated information is not a function of a specific pair of dose–response values, but rather it is an averaged quantity that depends on the probability of seeing any such random pair. The fundamental quantity in Shannon's theory is *entropy*, typically denoted S or H, which measures uncertainty in one's knowledge of a value of a variable [46]. It is measured in *bits*, the unit familiar to all of us in the digital age. One bit of uncertainty means that the variable can be completely specified by answering a single binary ('yes' or 'no') question. A measurement may not specify a variable of interest completely. For example, observing a response may still leave some uncertainty about the dose that has led to it. Then the amount of information that the response communicates about the dose, $I[r \to d]$, is the difference of the *a priori* dose entropy and its (averaged) entropy after observing the response. Interestingly, the law of multiplication of probabilities ensures that $I[r \to d] = I[d \to r]$. Thus the information is *mutual* [47], see Figure 1. This agrees with our observation that the inability to fully infer the dose from the response is equivalent to the inability to predict the response to the dose.

Mutual information depends on means and variances of distributions of cellular responses, but only because both can affect one's ability to infer the dose from the response. It can be calculated for any kind of variables, discrete, continuous, and multidimensional. Thus mutual information is a good candidate for a universal metric to characterize fidelity of biological communication in the presence of noise. Not surprisingly, it has been used widely in biology, achieving the largest impact in neurophysiology [52]. In molecular and cellular biology, information-theoretic approaches are still relatively rare. And yet they have been employed already to study processes as diverse as gene regulation, development, and protein signaling, both experimentally and computationally [53–55,56••,57••,58•,59•]. A good survey is [60].

Wider acceptance of mutual information for analysis of biological signaling is hindered by the fact that most practitioners have little intuitive understanding of how to interpret its values. For example, different authors measure mutual information of about 1 bit (between ~0.5 and ~2 bits) in a variety of cellular systems [54,56••,57••], but it is not immediately clear what this means. In fact, the interpretation is straightforward: a single bit allows to distinguish between two alternative possibilities. In the context of a dose response, if all doses are *a priori* equally likely, then 1 bit of information could mean that half of all dose values could not have caused a measured response. For example, with 1 bit, the presence or absence of a stimulant or drug can be distinguished with high fidelity from the response. More generally, a bit means that *a priori* uncertainty about the dose value is decreased in half, on average, following an observation of a response. If the number of bits is higher, the response resolution is exponentially greater. Three bits signify uncertainty reduction by a factor of $2^3 = 8$, or roughly down to one out of eight distinct levels of dose response; 4 bits correspond to $2^4 = 16$, levels, and so on, approaching infinite information for a perfect resolution of a one-to-one, deterministic dose–response. Fractional values of information are similarly interpretable: for example, information of 1.7 bits means that the response reduces the *a priori* uncertainly about the dose by $2^{1.7} \approx 3.25$, times. Thus in the simplest case, only about $\approx 1/3.25$ of *a priori* possible dose values could have led to an observed response.

The meaning of 1 bit is further revealed by noticing that the information between two normally distributed variables is $I = -(1/2)\log_2(1 - R^2)$, where R^2 is the usual

coefficient of determination (see e.g., [61•]). Thus, generally, mutual information can be viewed as a generalization of the coefficient of determination for nonlinear variable dependencies, where mutual information of 1, 2, 3, and 4 bits corresponds to the same amount of statistical dependence as is shared by normally distributed and linearly related variables with R^2 of 0.750, 0.938, 0.984, and 0.996, respectively. Thus on the one hand, the information of 1 bit is quite small since it allows to distinguish only two doses using response values. However, on the other hand, a single bit corresponds to an effective $R^2 = 0.750$, which is larger than resolution of typical biochemical experiments. This observation immediately suggests that measuring large information values requires experiments with physics-level precision, such as fluorescence tracking described earlier [56••]. Further, since information measures *all* dependences in data, it requires a lot larger data sets to be estimated reliably compared to simpler, linear measures of dependency [62–69], which makes experiments even harder.

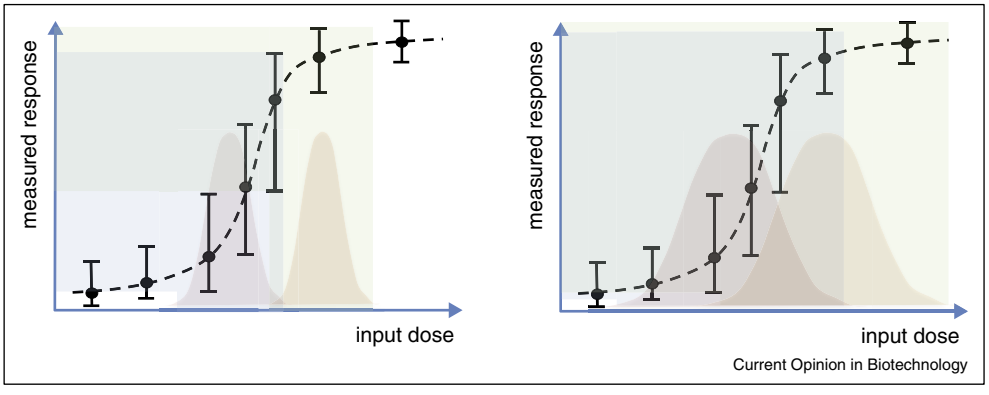
Such non-intuitive, exotic scaling of information values, and the difficulty in measuring them, make one ponder if characterizing the fidelity of a dose response in bits may be mathematically elegant, but ultimately not very useful. To disperse this concern, we now explore what using the mutual information may imply for our view of the noisy cellular processes.

Implications of measuring biological noise in bits

As noted above, in spite of sophisticated ways to measure cellular noise, the functional meaning of a specific noise magnitude in a given biological process is frequently unclear. Thus an important implication of using mutual information is that it provides such an *explicit interpretation*: noise limits the ability of the input to control the response of a system to the degree specified by the number of bits. Roughly speaking, information is the logarithm of the number of distinct, input-dependent states in the response repertoire of the cells. Conversely, the signal passing through a biochemical network is compromised by the noise, and the information is the number of distinguishable signal levels sensed by the cell.

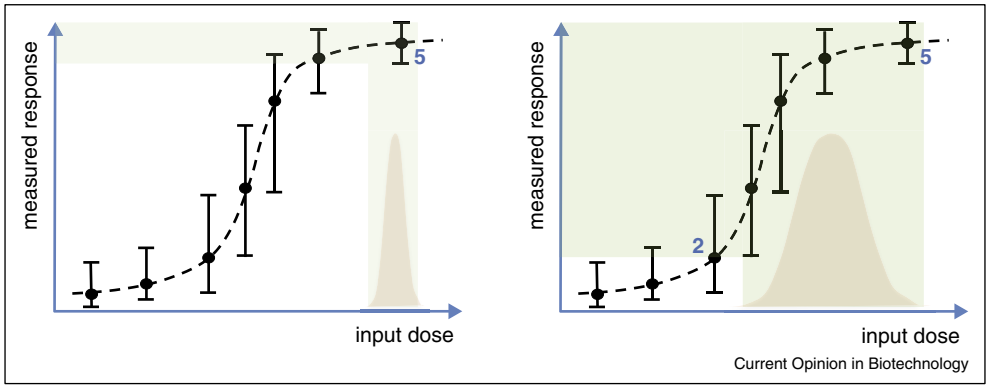
Changes in the information processing characteristics of a particular signaling network often signal a disease state. For instance, normal and cancer cells often have different responses to signaling molecules, such as growth factors, or to drugs. In fact, diseases such as cancer, which result from misregulation of cellular functions, can be interpreted as information diseases: a cancerous cell that proliferates irrespective of the concentration of an external growth factor transmits no information through the respective signaling pathway since the response (proliferation) is not informative of the signal (growth factor concentration). However, such defects in information-processing capabilities may be of different kinds. For example, average dose responses may be the same, but the cell may differ greatly in the variability of the cellular responses (cf. Figure 2). Alternatively, the average response characteristics, such as its dynamic range, or the threshold of sensitivity to the signal, might change in cancer versus normal cells without affecting the response noise levels. Traditional ways of quantifying properties of biochemical signaling systems make it hard to compare these cases in the same language. Indeed, viewing just the dose response curves, we cannot state if a twofold increase in the noise level, twofold decrease in the dynamics range, or a twofold change in the activation threshold

Figure 2



Many cellular diseases may be viewed as information diseases. Changes in means and the variability of dose response, as well as in the distribution of doses causes by diseases can result in a drastic reduction of the transmitted information. Here we illustrate just one possible effect. Consider a system that needs to be in one of two phenotypes depending on a value of some signaling molecule. In the left panel, the distributions of doses for these two phenotypes overlap only weakly. As a result, the responses to each case are quite distinct (blue and green shaded areas), and the cell has a high probability of responding with the right phenotype. In the right panel, the disease only increased the variability of doses for each of the two states. The distributions of responses overlap, the information transmitted about which of the two distributions the dose came from is low, and the cell will be choosing the response almost at random.

Figure 3



The mean and the error bars of the dose response are insufficient to characterize the information transduction capacity of the system, and the distribution of possible doses is also needed. (Left) Here the distribution of doses is very narrow (light red bell-shaped curve). Thus all possible responses are consistent with just one dose 5. Since no doses can be reliably distinguished, the information transduced in this system is close to zero. (Right) Here the distribution of observable doses spans a larger range, and the induced responses can lead to identifying doses anywhere between doses 2 and 5. Information in this system is much larger than in the left panel, even though the dose response is the same.

will have a larger effect on the signal transduction. Information-theoretic characterization of cellular signaling resolves this problem in a straightforward way.

Another important implication of using information to characterize noisy cellular signaling is that it explicitly forces us to consider *distributions of signals and responses*. Consider, for example, a typical saturating dose response with some dose-dependent noise encapsulated in experimental error bars, Figure 3. The information transmitted by this signaling system will depend on the distribution of possible doses [47]. If all possible doses correspond to the saturated mean response, or to responses with very large variances, then the information is low. In contrast, if the *a priori* probable doses span the entire dynamic range of the system, and probability of doses with low response variance is high, then the information is high as well, cf. Figure 3. Thus one cannot say whether noise is functionally important without saying first which doses the system is likely to experience. Correspondingly, in information-theoretic analysis, one often measures the natural distribution of doses and then controls it accurately experimentally (see e.g., [66,70–72] for some neurophysiology examples). Alternatively, one can ask which dose distribution would maximize information transmission through a given biological system [54,56••]. This puts the maximum limit on the amount of information that can be transmitted and also identifies which input values can deliver the most effect in terms of the response sensitivity.

One can take this analysis further and study which aspects of the input beyond its dose, such as the rate of its change or its duration, can affect the response the most [73,74,75•]. A greater amount of information contained in certain features of inputs suggests that these features can yield a greater degree of manipulation of the biochemical system, and are less likely to be affected by the noise. In the same vein, it is possible that a greater information could be delivered by multiple simultaneous signals, each with its own response and noise characteristics. Indeed, typical signaling systems participate in a variety of cellular decision-making processes, with multiple distinct response phenotypes. And yet experimental values of information of about 1 bit [56••] seem to suggest that only about two possible responses should be reliably distinguished. Thus richness of cellular behaviors requires either collective *cellular decision-making* or use of *complex, dynamic, combinatorial inputs* [56••,57••,76,77••], which can be identified with information-theoretic analysis. Even more interestingly, the type of the input yielding the best control of the cell population can dramatically change as a function of a disease state, suggesting new protocols that could maximize effects of drugs through more dynamically complex and combinatorial inputs.

Information-centric view of biological signaling affords an easier interpretation of data and identifiability of functionally important components. It is thus a natural framework for quantification of signaling in noisy biochemical networks. However, the approach allows a lot more. Indeed, high-fidelity information transmission allows making informed decisions, and is, therefore, a key contribution to an organism's fitness, likely to be optimized by evolution. Indeed, physiological adaptation optimizes transmitted information [52,61,70], signaling that various organisms 'care' about the transmitted bits. Even the very

topology of cellular regulatory networks may be optimized to allow more accurate information transmission [78]. Correspondingly, there is a series of theoretical arguments (most of which are not yet verified experimentally) that directly limit an organism's fitness by the amount of information the organism accumulates about its environment [42,79•,80]. This realization suggests that the next few years of research may show that information-theoretic approach is not one of the many possible, but rather *the only natural framework* for dealing with the single cell resolution analysis and the associated cell-cell variability in biochemical signaling.

Acknowledgements

This work was supported in part by James S. McDonnell foundation grant No. 220020321, HFSP Grant No. RGY0084/2011 (IN), and by NIH Grant No. GM072024 (AL).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Balazsi G, van Oudenaarden A, Collins JJ: **Cellular decision making and biological noise: from microbes to mammals**. *Cell* 2011, **144**:910-925.
2. Zhang J *et al.*: **Creating new fluorescent probes for cell biology**. *Nat Rev Mol Cell Biol* 2002, **3**:906-918.
3. Thattai M, van Oudenaarden A: **Intrinsic noise in gene regulatory networks**. *Proc Natl Acad Sci U S A* 2001, **98**:8614-8619.
4. Elowitz MB *et al.*: **Stochastic gene expression in a single cell**. *Science* 2002, **297**:1183-1186.
5. Raj A *et al.*: **Imaging individual mRNA molecules using multiple singly labeled probes**. *Nat Methods* 2008, **5**:877-879.
6. Raj A, van Oudenaarden A: **Nature, nurture, or chance: stochastic gene expression and its consequences**. *Cell* 2008, **135**:216-226.
7. Ozbudak EM *et al.*: **Multistability in the lactose utilization network of *Escherichia coli***. *Nature* 2004, **427**:737-740.
8. Rao CV, Wolf DM, Arkin AP: **Control, exploitation and tolerance of intracellular noise**. *Nature* 2002, **420**:231-237.
9. van Kampen NG: *Stochastic Processes in Physics and Chemistry*. edn 3. Amsterdam: North-Holland Personal Library; 2007, :. Boston: Elsevier. xvi, 463 p..
10. Gillespie D: **Exact stochastic simulation of coupled chemical reactions**. *J Phys Chem* 1977, **81**:2340.
11. Gillespie DT: **Stochastic simulation of chemical kinetics**. *Annu Rev Phys Chem* 2007, **58**:35-55.
12. Gibson M, Bruck J: **Efficient exact stochastic simulation of chemical systems with many species and many channels**. *J Phys Chem A* 2000, **104**:1876.
13. Cao Y, Gillespie D, Petzold L: **The slow-scale stochasitc simulation algorithm**. *J Chem Phys* 2005, **122**:014116.
14. Sinitsyn NA, Hengartner N, Nemenman I: **Adiabatic coarse-graining and simulations of stochastic biochemical networks**. *Proc Natl Acad Sci U S A* 2009, **106**:10546-10551.
15. Faeder JR, Blinov ML, Hlavacek WS: **Rule-based modeling of biochemical systems with BioNetGen**. *Methods Mol Biol* 2009, **500**:113-167.
16. Paulsson J: **Summing up the noise in gene networks**. *Nature* 2004, **427**:415-418.
17. Walczak AM, Mugler A, Wiggins CH: **A stochastic spectral analysis of transcriptional regulatory cascades**. *Proc Natl Acad Sci U S A* 2009, **106**:6529-6534.
18. Walczak AM, Mugler A, Wiggins CH: **Analytic methods for modeling stochastic regulatory networks**. *Methods Mol Biol* 2012, **880**:273-322.
19. Munsky B, Khammash M: **The finite state projection algorithm for the solution of the chemical master equation**. *J Chem Phys* 2006, **124**:044104.
20. Andrews SS, Bray D: **Stochastic simulation of chemical reactions with spatial resolution and single molecule detail**. *Phys Biol* 2004, **1**:137-151.
21. de Ronde W *et al.*: **Statistical properties of multistep enzyme-mediated reactions**. *IET Syst Biol* 2009, **3**:429.
22. Bel G, Munsky B, Nemenman I: **The simplicity of completion time distributions for common complex biochemical processes**. *Phys Biol* 2010, **7**:016003.
23. Sneddon MW, Faeder JR, Emonet T: **Efficient modeling, simulation and coarse-graining of biological complexity with NFsim**. *Nat Methods* 2011, **8**:177-183.
24. Machta BB *et al.*: **Parameter space compression underlies emergent theories and predictive models**. *Science* 2013, **342**:604-607.
25. Swain PS, Elowitz MB, Siggia ED: **Intrinsic and extrinsic contributions to stochasticity in gene expression**. *Proc Natl Acad Sci U S A* 2002, **99**:12795-12800.
26. Raser JM, O'Shea EK: **Control of stochasticity in eukaryotic gene expression**. *Science* 2004, **304**:1811-1814.
27. Ozbudak EM *et al.*: **Regulation of noise in the expression of a single gene**. *Nat Genet* 2002, **31**:69-73.
28. Pedraza JM, van Oudenaarden A: **Noise propagation in gene networks**. *Science* 2005, **307**:1965-1969.
29. Suel GM *et al.*: **An excitable gene regulatory circuit induces transient cellular differentiation**. *Nature* 2006, **440**:545-550.
30. Paliwal S *et al.*: **MAPK-mediated bimodal gene expression and adaptive gradient sensing in yeast**. *Nature* 2007, **446**:46-51.
31. Cagatay T *et al.*: **Architecture-dependent noise discriminates functionally analogous differentiation circuits**. *Cell* 2009, **139**:512-522.
32. Ferrell JE Jr, Machleder EM: **The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes**. *Science* 1998, **280**:895-898.
33. Mettetal JT *et al.*: **Predicting stochastic gene expression dynamics in single cells**. *Proc Natl Acad Sci U S A* 2006, **103**:7304-7309.
34. Munsky B, Neuert G, van Oudenaarden A: **Using gene expression noise to understand gene regulation**. *Science* 2012, **336**:183-187.
35. • Neuert G *et al.*: **Systematic identification of signal-activated stochastic gene regulation**. *Science* 2013, **339**:584-587.
This paper is a great illustration of the precision with which cellular noises can be measured and modeled. In fact, here the properties of the noise were used to infer the structure of the regulatory network that responds to osmotic shock in yeast.
36. Kaern M *et al.*: **Stochasticity in gene expression: from theories to phenotypes**. *Nat Rev Genet* 2005, **6**:451-464.
37. Eldar A, Elowitz MB: **Functional roles for noise in genetic circuits**. *Nature* 2010, **467**:167-173.
38. Balaban NQ *et al.*: **Bacterial persistence as a phenotypic switch**. *Science* 2004, **305**:1622-1625.
39. • Deris JB *et al.*: **The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria**. *Science* 2013, **342**:1237435.
This paper argues that the noise in cellular regulation, amplified through a positive feedback loop, can result in a drastic population heterogeneity, which in its turn can lead to antibiotic persistence in a population. Unlike in

a prevailing paradigm, here heterogeneity is generated in response to the environment change.

40. Gallistel CR, Fairhurst S, Balsam P: **The learning curve: implications of a quantitative analysis**. *Proc Natl Acad Sci U S A* 2004, **101**:13124-13131.

41. Kussell E *et al.*: **Bacterial persistence: a model of survival in changing environments**. *Genetics* 2005, **169**:1807-1814.

42. Kussell E, Leibler S: **Phenotypic diversity, population growth, and information in fluctuating environments**. *Science* 2005, **309**:2075-2078.

43. Lewis K: **Persister cells**. *Annu Rev Microbiol* 2010, **64**:357-372.

44. Johnson PJ, Levin BR: **Pharmacodynamics, population dynamics, and the evolution of persistence in *Staphylococcus aureus***. *PLoS Genet* 2013, **9**:e1003123.

45. Dean M, Fojo T, Bates S: **Tumour stem cells and drug resistance**. *Nat Rev Cancer* 2005, **5**:275-284.

46. Shannon CE, Weaver W: *The Mathematical Theory of Communication*. Urbana: University of Illinois Press; 1949, .

47. Cover TM, Thomas JA: *Elements of Information Theory*. New York: John Wiley & Sons; 1991, .

48. Mackay DJC: *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press; 2003, .

49. El Gamal AA, Kim Y-H: *Network Information Theory*. Cambridge, New York: Cambridge University Press; 2011, :. xxviii, 685 p..

50. Schneidman E *et al.*: **Network information and connected correlations**. *Phys Rev Lett* 2003, **91**:238701.

51. Margolin AA *et al.*: **Multivariate dependence and genetic networks inference**. *IET Syst Biol* 2010, **4**:428.

52. Fairhall A, Shea-Brown E, Barreiro A: **Information theoretic approaches to understanding circuit function**. *Curr Opin Neurobiol* 2012, **22**:653-659.

53. Ziv E, Nemenman I, Wiggins CH: **Optimal signal processing in small stochastic biochemical networks**. *PLoS ONE* 2007, **2**:e1077.

54. Tkacik G, Callan CG Jr, Bialek W: **Information flow and optimization in transcriptional regulation**. *Proc Natl Acad Sci U S A* 2008, **105**:12265-12270.

55. Tostevin F, ten Wolde PR: **Mutual information between input and output trajectories of biochemical networks**. *Phys Rev Lett* 2009, **102**:218101.

56. •• Cheong R *et al.*: **Information transduction capacity of noisy biochemical signaling networks**. *Science* 2011, **334**:354-358.
This is one of the first papers that analyzed experimental cellular signal transduction data in the language of information theory. It has touched on many of the topics raised in this opinion: dependence of the information on the distribution of environment, difficulty in measuring information, using information to understand which parts of the network dominate errors in signal transduction, and using combinatorial, collective, and dynamical encoding of external signals.

57. •• Dubuis JO *et al.*: **Positional information, in bits**. *Proc Natl Acad Sci U S A* 2013, **110**:16301-16308.
Through combination of precise experiments and mathematical modeling, this paper analyzes combinatorial encoding of positional information in the early development of a fly embryo.

58. • Mugler A, Tostevin F, ten Wolde PR: **Spatial partitioning improves the reliability of biochemical signaling**. *Proc Natl Acad Sci U S A* 2013, **110**:5927-5932.
This paper argues that spatial inhomogeneity of distribution of signaling molecules inside a cell can result in higher information transmission than a well-mixed scenario.

59. • Uda S *et al.*: **Robustness and compensation of information transmission of signaling pathways**. *Science* 2013, **341**:558-561.
The main contribution of this paper is in pointing out that the signal transduction in cellular networks, as quantified by information-theoretic quantities, is more robust to various signal and network perturbations than simpler measures of quality of signal transduction; cells compensate for such perturbations to achieve stable information transmission.

60. Tkacik G, Walczak AM: **Information transmission in genetic regulatory networks: a review**. *J Phys Condens Matter* 2011, **23**:p153102.

61. • Nemenman I: **Information theory and adaptation**. In *Quantitative Biology: From Molecules to Cellular Systems.* Edited by Wall M. CRC Press; 2012.
This is a review of information-theoretic approaches in biology, which relies a bit more on mathematical formalism than the current opinion. Nonetheless, it is accessible to the audience without specialized mathematical background.

62. Strong SP *et al.*: **Entropy and information in neural spike trains**. *Phys Rev Lett* 1998, **80**:197.

63. Nemenman I, Shafee F, Bialek W: **Entropy and inference, revisited**. In *Advances in Neural Information Processing Systems*, vol 14. Edited by Dietterich TG, Becker S, Ghahramani Z. MIT Press; 2002.

64. Paninski L: **Estimation of entropy and mutual information**. *Neur Comp* 2003, **15**:1191-1253.

65. Kraskov A, Stoegbauer H, Grassberger P: **Estimating mutual information**. *Phys Rev E* 2004, **69**:066138.

66. Nemenman I *et al.*: **Neural coding of natural stimuli: information at sub-millisecond resolution**. *PLoS Comput Biol* 2008, **4**:e1000025.

67. Nemenman I: **Coincidences and estimation of entropies of random variables with large cardinalities**. *Entropy* 2011, **13**:2013-2023.

68. Slonim N *et al.*: **Information-based clustering**. *Proc Natl Acad Sci U S A* 2005, **102**:18297-18302.

69. Grassberger P: *Entropy Estimates from Insufficient Samplings*. 2003arXiv:physics/0307138.

70. Laughlin S: **A simple coding procedure enhances a neuron's information capacity**. *Z Naturforsch C* 1981, **36**:910-912.

71. Reinagel P, Laughlin S: **Natural stimulus statistics**. *Network* 2001, **12**:237-240.

72. Tkacik G *et al.*: **Natural images from the birthplace of the human eye**. *PLoS ONE* 2011, **6**:e20409.

73. Tishby N, Pereira FC, Bialek W: **The information bottleneck method**. In *Proceedings of 37th Annual Allerton Conference on Communication, Control and Computing*. 1999.

74. Sharpee T, Rust NC, Bialek W: **Analyzing neural responses to natural signals: maximally informative dimensions**. *Neural Comput* 2004, **16**:223-250.

75. • Stromberg SP, Antia R, Nemenman I: **Population-expression models of immune response**. *Phys Biol* 2013, **10**:035010.
This computational paper introduces mathematical tools for simultaneous analysis of intrinsic cellular stochasticity and the variability in phenotypic decisions performed by the cells in a population. It further proposes computational methods, some of them based on information-theoretic approaches, to find the dynamical variables characterizing the cellular noise that are the most functionally important.

76. Buchler NE, Gerland U, Hwa T: **On schemes of combinatorial transcription logic**. *Proc Natl Acad Sci U S A* 2003, **100**:5136-5141.

77. •• Toettcher JE, Weiner OD, Lim WA: **Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module**. *Cell* 2013, **155**:1422-1434.
This work uses optogenetics to investigate reproducibility of dose–response curves within a single cell versus in cellular populations. It further looks at dynamic encoding of information. However, it remains unclear if higher intrinsic versus population reproducibility can be used to sculpt a reliable population response.

78. Walczak AM, Tkacik G, Bialek W: **Optimizing information flow in small genetic networks. II. Feed-forward interactions**. *Phys Rev E Stat Nonlin Soft Matter Phys* 2010, **81(Pt 1)**:041905.

79. Rivoire O, Leibler S: **The value of information for populations in varying environments**. *J Stat Phys* 2011, **142**:1124-1166.
• This is a capstone paper in a series of theoretical investigations that shows a tight link between the information an individual in a heterogeneous population accumulates about its environment and the rate of growth of such populations, under various assumptions about the structure of the environment, the population, and the individual decision-making strategies.

80. Bergstrom C, Lachmann M: **Shannon information and biological fitness**. *IEEE Information Theory Workshop*. 2004.