

RESEARCH ARTICLE

# Simple biochemical networks allow accurate sensing of multiple ligands with a single receptor

Vijay Singh<sup>1,2</sup>, Ilya Nemenman<sup>1,3,4\*</sup>

**1** Department of Physics, Emory University, Atlanta, Georgia, United States of America, **2** Computational Neuroscience Initiative, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America, **3** Department of Biology, Emory University, Atlanta, Georgia, United States of America, **4** Initiative in Theory and Modeling of Living Systems, Emory University, Atlanta, Georgia, United States of America

\* [ilya.nemenman@emory.edu](mailto:ilya.nemenman@emory.edu)



**OPEN ACCESS**

**Citation:** Singh V, Nemenman I (2017) Simple biochemical networks allow accurate sensing of multiple ligands with a single receptor. *PLoS Comput Biol* 13(4): e1005490. <https://doi.org/10.1371/journal.pcbi.1005490>

**Editor:** Alexandre V Morozov, Rutgers University, UNITED STATES

**Received:** May 20, 2016

**Accepted:** March 31, 2017

**Published:** April 14, 2017

**Copyright:** © 2017 Singh, Nemenman. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper.

**Funding:** This work was supported in part by the James S. McDonnell Foundation grant 220020321 and by the National Science Foundation grants IOS-1208126 and PoLS-1410978. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Cells use surface receptors to estimate concentrations of external ligands. Limits on the accuracy of such estimations have been well studied for pairs of ligand and receptor species. However, the environment typically contains many ligands, which can bind to the same receptors with different affinities, resulting in cross-talk. In traditional rate models, such cross-talk prevents accurate inference of concentrations of individual ligands. In contrast, here we show that knowing the precise timing sequence of stochastic binding and unbinding events allows one receptor to provide information about multiple ligands simultaneously and with a high accuracy. We show that such high-accuracy estimation of multiple concentrations can be realized with simple structural modifications of the familiar kinetic proofreading biochemical network diagram. We give two specific examples of such modifications. We argue that structural and functional features of real cellular biochemical sensory networks in immune cells, such as feedforward and feedback loops or ligand antagonism, sometimes can be understood as solutions to the accurate multi-ligand estimation problem.

## Author summary

Cells live in chemically complex environments with many different chemical ligands around them. Can cells estimate concentrations of more ligands than they have receptor types? In this paper, we show that, surprisingly, the answer is “yes”, and the estimation can be implemented with simple biochemical components already present in many cells. Therefore, cells may “know” a lot more about their environment and thus may be able to implement more complex and accurate response strategies than was previously thought.

## Introduction

Cells obtain information about their environment by capturing ligand molecules with receptors on their surface and estimating the ligand concentration from the receptor activity. Limits

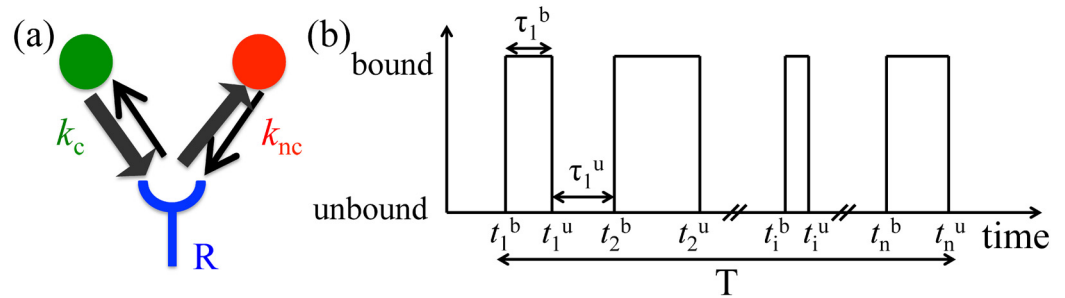
on the accuracy of such estimation have been a subject of interest since the seminal work of Berg and Purcell [1], with several substantial extensions found recently [2–8]. Most of these assume one ligand species coupled to one receptor species, and the actual detection in most of these models is rather simple, involving counting the number or the duration of binding / unbinding events over a specific period of time.

However, cells carry many types of receptors and have many species of ligands around them. The same ligand can bind to many receptors, albeit with different affinities, and vice versa. This is commonly referred to as *cross-talk*. At the same time, real cellular sensory systems are incredibly complex, involving many dozens of identified biochemical species downstream of a typical receptor [9]. Functionally many of such signaling motifs are probably related to solving the cross-talk problem [10, 11], and are a topic of active research.

In traditional deterministic chemical kinetics, one cannot estimate concentrations of more ligands than there are receptor types. Further, even a weak cross-talk prevents determination of concentrations of individual chemical species since the activity of a receptor is a function of a weighted sum of concentrations of all ligands that can bind to it. In contrast, here we argue that, with cross-talk, concentration of more than one chemical species can be inferred from the activity of one receptor, provided that the stochastic temporal sequence of receptor binding and unbinding events is accessible instead of its mean occupancy. This is an important departure from the traditional view of cellular signaling that posits as many receptor types as there are ligand concentrations to be estimated. Indeed, previous works studying temporal sequences of receptor occupancy for ligand detection [11] and concentration estimation [5, 13, 12] have only considered the detection/estimation of a single ligand present in a mixture. We argue that the receptor occupancy sequence contains much more information about the mixture. In fact, based on the maximum likelihood techniques, which have been used previously to study receptor occupancy, we show that *all* components of the ligand mixture can be estimated by just one receptor, at least in principle. This surprising result can be understood by noting that a typical duration of time that a ligand remains bound to the receptors depends on its unbinding rate. Thus observing the statistics of the receptor's *unbound* time durations allows estimation of a weighted average of all chemical species that interact with it [5]. Then the statistics of the *bound* time durations tells how common each ligand is.

The result is very general and independent on the choice of a downstream biochemical kinetics scheme that actually performs the estimation. In this article, we derive the result for the simplest problem of this class, namely one receptor interacting with two ligand species. While the exact solution of the inference problem for finding both ligand concentrations is hard to implement using common biochemical machinery, we show that an accurate approximation is possible using simple extensions of the familiar kinetic proofreading mechanism [14, 15]. We identify examples of such motifs implementing such estimation of multiple concentrations in signaling networks found downstream of many immune receptors [9], arguing that real biological systems may be implementing such multivariate concentration sensing. The kinetic schemes that we analyze detect rare ligands more accurately than a simple kinetic proofreading does, and we argue that the involved biochemical computation can explain properties like ligand antagonism, commonly observed in receptor signaling.

Overall, these different arguments support our main idea, that *the temporal sequence of binding and unbinding on a single receptor can provide an accurate estimate of the concentration of multiple ligands that bind to the receptor, and that the involved calculations can be performed reliably by known biochemical networks.*



**Fig 1. The model.** (a). Two ligands, cognate and non-cognate having concentrations  $c_c$  and  $c_{nc}$ , bind to a receptor R with binding rates  $k_c$  and  $k_{nc}$ , respectively. The cognate unbinding rate is defined as lower than the non-cognate one ( $r_c < r_{nc}$ ). (b) Time series of receptor occupancy is used to determine both on-rates.

<https://doi.org/10.1371/journal.pcbi.1005490.g001>

## Results

### The model

Consider a single receptor interacting with a cognate and a non-cognate ligand (Fig 1) that have the concentrations  $c_c$  and  $c_{nc}$ , respectively. The binding rate of the ligands to the receptor are  $k_c$  and  $k_{nc}$ . The binding rates are diffusion limited and hence  $k_c \sim k_{nc}$ . It is the unbinding or off-rates,  $r_c$  and  $r_{nc}$ , that distinguish the two ligands:  $r_{nc} > r_c$ , and a cognate molecule typically stays bound for longer. The binding and unbinding rates ( $k_\alpha$ 's and  $r_\alpha$ 's) are fixed and can be assumed known for each receptor-ligand pair. Thus we are interested in the estimation of the ligand concentrations only,  $c_c$  and  $c_{nc}$ . Following Ref. [5], we estimate  $c_c$  and  $c_{nc}$  from the time-series of binding,  $\{t_i^b\}$ , and unbinding,  $\{t_i^u\}$ , events of a total duration  $T$  using Maximum Likelihood techniques, paralleling a recent similar independent discussion, which focused on detection of a single ligand concentration [12]. The numbers of binding and unbinding events are different by, at most, one, which is insignificant since we consider  $T \rightarrow \infty$ . Thus without loss of generality, we assume that the first event was a binding event at  $t_1^b$ , and the last one was the unbinding at  $t_n^u$ . We write the probability distribution of observing the sequence  $\{t_1^b, t_1^u, \dots, t_n^b, t_n^u\}$ , or alternatively the sequence of binding and unbinding intervals  $\tau_i^b = t_i^u - t_i^b$ , and  $\tau_i^u = t_{i+1}^b - t_i^u$ :

$$P \equiv P(\{\tau_i^b, \tau_i^u\} | c_c, c_{nc}) = \frac{1}{Z} \prod_{i=1}^n \left[ e^{-\tau_i^u (k_c c_c + k_{nc} c_{nc})} \left( k_c c_c r_c e^{-\tau_i^b r_c} + k_{nc} c_{nc} r_{nc} e^{-\tau_i^b r_{nc}} \right) \right]. \quad (1)$$

Here the first term under the product sign is the probability of the receptor staying unbound for  $\tau_i^u$ . The second term, which from now on we denote by  $D(c_c, c_{nc}, \tau_i^b)$ , is proportional to the probability of staying bound for  $\tau_i^b$ .  $D(c_c, c_{nc}, \tau_i^b)$  has contributions from binding events from both the cognate and the noncognate ligands, with odds of  $c_c$  and  $c_{nc}$ , respectively. Finally,  $Z$  is the normalization,

$$Z = \sum P(\{\tau_i^b, \tau_i^u\} | c_c, c_{nc}), \quad (2)$$

where the sum is over all sequences of duration  $T$  and  $n$  binding-unbinding events. Note that here we define  $\tau_n^u = t_1^b + (T - t_n^u)$ , so that the  $n$ 'th unbound interval includes the "incomplete" unbound intervals before the first binding and after the last unbinding.

The log-likelihood of  $c_c$  and  $c_{nc}$  is the logarithm of  $P$ , Eq (1). Taking the derivatives of the log-likelihood w. r. t.  $c_c$  and  $c_{nc}$  and setting them to zero gives the Maximum Likelihood (ML) equations for the concentrations. Denoting by  $T^u = \sum_{i=1}^n \tau_i^u$ , the total time the receptor is

unbound, these ML equations are (see [Methods](#) for the derivation):

$$-k_c T^u + \sum_{i=1}^n \frac{k_c r_c e^{-\tau_i^b r_c}}{D(c_c^*, c_{nc}^*, \tau_i^b)} = 0, \tag{3}$$

$$-k_{nc} T^u + \sum_{i=1}^n \frac{k_{nc} r_{nc} e^{-\tau_i^b r_{nc}}}{D(c_c^*, c_{nc}^*, \tau_i^b)} = 0, \tag{4}$$

where  $c_c^*$  and  $c_{nc}^*$  denotes the ML solution. Multiplying Eqs (3) and (4) by  $c_c^*$  and  $c_{nc}^*$ , respectively, and adding them gives

$$k_c c_c^* + k_{nc} c_{nc}^* = \frac{n}{T^u}. \tag{5}$$

As in Ref. [5], the total on-rate (the weighted average of the external concentrations) is determined only by the average duration of the unbound interval,  $(n/T^u)^{-1}$ , because no binding is possible when the receptor is already bound. For the special case of  $k_c \approx k_{nc} \approx k$  (for ligands with binding rate determined by diffusion), Eq (5) determines the maximum likelihood estimate of the sum of the two concentrations, similar to the result in Ref. [5, 12]:

$$c_{tot}^* = (c_c^* + c_{nc}^*) = \frac{n}{T^u k}. \tag{6}$$

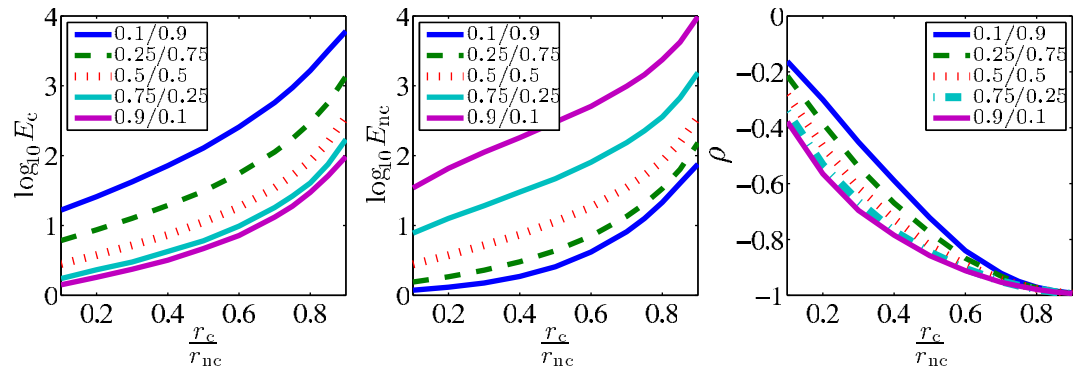
This shows that the estimates are negatively correlated. For general  $k_i$ 's, a weighted sum of the concentrations is determined, but the negative correlation persists.

To get the individual concentrations, we need to solve the ML equations Eqs (3) and (4). In general, they can only be solved numerically. However, as all ML estimators, they are unbiased to the leading order in  $n$  (we verified this numerically). The standard errors of the ML estimates can be obtained by inverting the Hessian matrix,

$$\left. \frac{\partial^2 \log P}{\partial c_\alpha \partial c_\beta} \right|_{c_c^*, c_{nc}^*} = \sum_{i=1}^n \left[ \frac{-1}{D(c_c, c_{nc}, \tau_i^b)^2} \times \begin{pmatrix} k_c^2 r_c^2 e^{-2\tau_i^b r_c} & k_c k_{nc} r_c r_{nc} e^{-\tau_i^b (r_c + r_{nc})} \\ k_c k_{nc} r_c r_{nc} e^{-\tau_i^b (r_c + r_{nc})} & k_{nc}^2 r_{nc}^2 e^{-2\tau_i^b r_{nc}} \end{pmatrix} \right], \tag{7}$$

where greek indices stand for  $\{c, nc\}$ . Each term in the Hessian matrix is a sum of  $n$  numbers, each smaller than zero. The inverse of  $\frac{\partial^2 \log P}{\partial c_\alpha \partial c_\beta}$ , which scales as  $\propto 1/n$ , sets the minimum variance of any unbiased estimator according to the Cramer-Rao bound. It has straightforward analytical approximations in various regimes. For example, when the noncognate ligand is almost absent ( $c_c/c_{nc} \gg 1$ ), and its few molecules do not bind for long ( $r_c/r_{nc} \ll 1$ ), one gets  $\sigma^2(c_c^*) \approx (\partial^2 \log P / \partial c_c^2)_{c_c=c_c^*}^{-1} \approx 1/n$ , matching the accuracy of sensing one ligand with one receptor [5]. A regime relevant for detection of a rare, but highly specific ligand [11, 12, 16] can be investigated as well. For now, we focus on how the receptor estimates (rather than detects) concentrations of *both* ligands simultaneously, which requires us to explore the full range of on- and off-rates.

The estimates of the concentration  $c_c$  and  $c_{nc}$  are obtained by numerically solving ML equations, Eqs (3) and (4). We study the variability of these ML estimators in terms of their posterior variances. Notice that these posterior variances scale as  $1/n$ , so we define the error of the ML estimators,  $E$ , as the squared coefficient of variation times the number of binding-unbinding events,  $n$ . Hence, we have,  $E_c = n\sigma^2(c_c^*)/c_c^2$  and  $E_{nc} = n\sigma^2(c_{nc}^*)/c_{nc}^2$  for cognate and non-cognate ligands, respectively. These quantities have a finite limit at  $n \rightarrow \infty$ . Specifically,  $E = 1$  is the accuracy that a receptor that binds only a single ligand can obtain [5]. Thus  $E_c$  and  $E_{nc}$



**Fig 2. Variability of the ML estimators, represented by  $\log_{10} E_c$  (left),  $\log_{10} E_{nc}$  (center), and the correlation coefficient  $\rho$  between  $c_c^*$  and  $c_{nc}^*$  (right) as functions of the ratio of unbinding rates.** Here we use  $r_{nc} = 1$ ,  $k_c = k_{nc} = 1$ ,  $c_c + c_{nc} = 1$ . The legend and the colors represent different ratios of concentrations of the cognate and the non-cognate ligands ( $\frac{c_c}{c_{nc}}$ ). We plot averages over 30,000 randomly generated binding/unbinding sequences for each combination of the rates. Each sequence itself consists of  $n = 30,000$  binding events, simulated using the Gillespie algorithm. Standard errors are too small to show.

<https://doi.org/10.1371/journal.pcbi.1005490.g002>

compare the performance of our multi-ligand ML estimator to the limit achievable by a single ligand ML estimator. We show  $\log_{10} E_c$  and  $\log_{10} E_{nc}$  for different concentrations and off-rates in Fig 2. If the two ligands are readily distinguishable,  $r_c \ll r_{nc}$ , then the ligand with the larger concentration has  $E \sim 1$ . When  $c_c \sim c_{nc}$ ,  $E_i \sim 4 \dots 5$ , and it grows to  $10 \dots 30$  for a ligand with a very small relative concentration. Emphasizing the importance of the time scale separation,  $E > 100$  if the ligands are hard to distinguish,  $r_c \sim r_{nc}$ . Here the correlation coefficient  $\rho$  of the two estimates reaches  $-1$  because the same binding event can be attributed to either ligand. Finally, the asymmetry of the plots w. r. t. the exchange of  $c_c$  and  $c_{nc}$  is because the cognate ligand can generate short binding events, while long events from the noncognate ligand are exponentially unlikely. In summary, it is possible to infer two ligand concentrations from one receptor, with the error of only  $1 \dots 10$  times larger than for ligand-receptor pairs with no cross talk, as long as the two off-rates are substantially different. This complements the findings of [12] that a single concentration can be inferred from a time series of “on” and “off” events in a background of noncognate bindings using Maximum Likelihood estimation. We have verified that the analytical expression for the estimation error derived in Ref. [12] for a single cognate ligand matches our numerical results (see Methods).

### Approximate solution

It is not clear if there exist biochemical networks that can solve the ML equations, Eqs (3) and (4), exactly. Luckily, an approximate solution exists. Note that most of the long binding events come from the cognate ligand since the noncognate one dissociates faster. Defining long events as  $\tau_i^b \geq T^c$  and using Eq (5), we rewrite Eq (3) as

$$\frac{k_c n}{k_c c_c^* + k_{nc} c_{nc}^*} = \left( \sum_{\tau_i^b \geq T^c} + \sum_{\tau_i^b < T^c} \right) \frac{k_c r_c e^{-\tau_i^b r_c}}{D(c_c^*, c_{nc}^*, \tau_i^b)} \tag{8}$$

Assuming that all long events are cognate,  $T^c \gg 1/r_{nc}$ , gives

$$\frac{k_c n}{(k_c c_c^a + k_{nc} c_{nc}^a)} = \frac{n_l}{c_c^a} + \sum_{\tau_i^b < T^c} \frac{k_c r_c e^{-\tau_i^b r_c}}{D(c_c^a, c_{nc}^a, \tau_i^b)}, \tag{9}$$

where  $n_1$  is the number of long events, and the superscript “a” stands for the approximate solution. If further  $T$  is long enough so that there are many short events, and a single binding duration hardly affects  $k_c^*$ , then the sum in Eq (9) can be approximated by the expectation value:

$$\frac{n}{(k_c c_c^a + k_{nc} c_{nc}^a)} = \frac{n_1}{k_c c_c^a} + (n - n_1) \int_0^{T^c} \frac{r_c e^{-\tau^b r_c} P(\tau^b | c_c^a, c_{nc}^a) d\tau^b}{D(c_c^a, c_{nc}^a, \tau^b)}, \tag{10}$$

where  $P(\tau^b | c_c^a, c_{nc}^a)$  is the probability of observing a binding event of duration  $\tau^b$  for the given binding rates,

$$P(\tau^b | c_c^a, c_{nc}^a) = \frac{D(c_c^a, c_{nc}^a, \tau^b)}{(k_c c_c^a + k_{nc} c_{nc}^a)}. \tag{11}$$

Plugging Eq (11) into Eq (10), we obtain

$$\frac{1}{(k_c c_c^a + k_{nc} c_{nc}^a)} = \frac{n_1}{n k_c c_c^a} + \left(1 - \frac{n_1}{n}\right) \frac{1 - e^{-r_c T^c}}{(k_c c_c^a + k_{nc} c_{nc}^a)}. \tag{12}$$

Finally, since  $n_1 \ll n$ , using Eq (5), we get (see Methods for a detailed derivation):

$$c_c^a = \frac{1}{k_c} \frac{n_1}{T^u} e^{r_c T^c}, \tag{13}$$

$$c_{nc}^a = \frac{1}{k_{nc}} \left( \frac{n}{T^u} - \frac{n_1}{T^u} e^{r_c T^c} \right). \tag{14}$$

In other words, the approximate cognate ligand concentration is proportional to the number of long events.

We can estimate the bias and the variance of  $c_c^a$  and  $c_{nc}^a$  in a limiting case. If  $r_c$  and  $r_{nc}$  are not very different from each other, then one needs to focus on extremely long events in order to identify cognate bindings. This is only possible if  $T^c$  is much larger than the inverse of both of the unbinding rates,  $T^c \gg \{r_{nc}^{-1}, r_c^{-1}\}$ . Large  $T^c$  ensures that the long binding events get no or minimal contribution from non-cognate ligands. However, since the time for which the receptor stays bound is exponentially distributed, under this condition, the number of “long” events (such that  $\tau_b > T^c$ ) would be very small,  $n_1 \ll n$ . Thus most of the variance of  $c_c^a$  and  $c_{nc}^a$  in Eqs (13) and (14) comes from the variability of  $n_1$ , but not  $T^u$  (since  $T^u \propto n$ ). Thus we write  $\langle c_c^a \rangle \approx \frac{\langle n_1 \rangle e^{r_c T^c}}{\langle T^u \rangle k_c}$ . Further, the individual unbound periods are independent, so that  $\langle T^u \rangle = n \langle \tau^u \rangle = n / (k_c c_c + k_{nc} c_{nc})$  (notice the use of  $c$  rather than  $c^a$  here). Further,  $\langle n_1 \rangle = n P(\tau^b > T^c) = \frac{n}{(k_c c_c + k_{nc} c_{nc})} (k_c c_c e^{-r_c T^c} + k_{nc} c_{nc} e^{-r_{nc} T^c})$ . Combining these expressions, we get

$$\langle c_c^a \rangle \approx c_c + \frac{k_{nc} c_{nc}}{k_c} e^{-(r_{nc} - r_c) T^c}. \tag{15}$$

Thus for large  $T^c$ , the bias of the approximate estimator,  $\frac{k_{nc} c_{nc}}{k_c} e^{-(r_{nc} - r_c) T^c}$ , grows with the relative number of noncognate long bindings events. In turn, the latter is proportional to  $c_{nc}$ , but decreases exponentially with  $T^c$ .

Within the same approximation, the variance of the estimator is given by  $\sigma^2(c_c^a) \approx \frac{\sigma^2(n_1) e^{2r_c T^c}}{\langle T^u \rangle^2 k_c^2}$ . However, long binding events are rare, independent of each other, and

hence obey the Poisson statistics. Thus  $\sigma^2(n_i) = \langle n_i \rangle$ , so that

$$\sigma^2(c_c^a) \approx \langle c_c^a \rangle \frac{(c_c + k_{nc}c_{nc}/k_c)}{n} e^{r_c T^c}. \tag{16}$$

The variance obviously grows with  $T^c$ .

Knowing that the bias and the variance of the approximation change in opposite directions with  $T^c$ , we can find the optimal cutoff ( $T^c_*$ ) by minimizing the overall error. We define such error  $L$  as the sum of the variance and the squared bias of the estimator, i. e.,

$$L_c = \left[ \sigma^2(c_c^a) + (c_c - \langle c_c^a \rangle)^2 \right], \tag{17}$$

$$L_{nc} = \left[ \sigma^2(c_{nc}^a) + (c_{nc} - \langle c_{nc}^a \rangle)^2 \right]. \tag{18}$$

The optimal cutoff is obtained by minimizing  $L_c$  or, in other words, solving the bias-variance tradeoff:

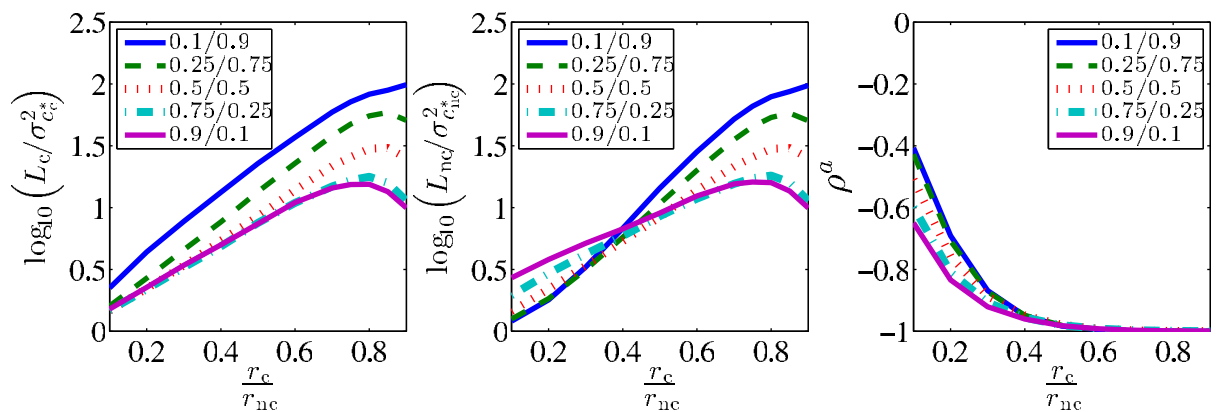
$$T^c_* = \arg \min_{T^c} L_c. \tag{19}$$

Near the optimal cutoff, the bias is small, and we use  $c_c$  instead of  $c_c^a$  for the variance of the estimator, Eq (16). Then solving Eq (19) gives:

$$T^c_* = \frac{1}{(2r_{nc} - r_c)} \log \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right]. \tag{20}$$

Plugging this into Eqs (15) and (16), we get the minimal error of the estimator, which we omit here for brevity.

The optimal cutoff is proportional to  $1/r_{nc}$  if  $r_{nc} \gg r_c$ , and it grows with  $r_c$ , allowing for better disambiguation of cognate and noncognate events. Crucially, the off-rates are dictated by the ligand identities. In contrast, the concentrations,  $c_c$  and  $c_{nc}$ , are what the receptors measures. Therefore, it is encouraging that  $T^c_*$  depends only logarithmically on the concentrations (and also on the duration of the measurement,  $T^u$ ). Thus even if  $T^c$  is fixed as  $T^c_*$  at some fixed values of  $c_c, c_{nc}$ , it remains near-optimal for a broad range of external concentrations. To illustrate this, we use  $T^c = T^c_*(k_c c_c = k_{nc} c_{nc} = 1/2) \equiv T_0$  and analyze the quality of the



**Fig 3. Comparison of errors of the approximate and the ML solutions.** We plot  $\log_{10}(L_c(T_0)/\sigma_{c_c}^2)$  (left),  $\log_{10}(L_{nc}(T_0)/\sigma_{c_{nc}}^2)$  (center) and the covariance of the approximate estimates (right) as functions of on- and off-rates. Simulations are performed in the same way as in Fig 2. Legends and color scheme are the same as in Fig 2.

<https://doi.org/10.1371/journal.pcbi.1005490.g003>



approximation in Fig 3, where we plot the ratio  $L_c(T_0)/\sigma_{c_c}^2$  and  $L_{nc}(T_0)/\sigma_{c_{nc}}^2$ . Notice that  $\sigma_{c_c}^2$  and  $\sigma_{c_{nc}}^2$ , the variances of the exact ML estimators, are proportional to  $E_c$  and  $E_{nc}$ , respectively. Since ML estimators are unbiased, the ratios  $L_c(T_0)/\sigma_{c_c}^2$  and  $L_{nc}(T_0)/\sigma_{c_{nc}}^2$  compare the errors of the approximate solution to the errors  $E_c$  and  $E_{nc}$ . Since these ratios approach 1 when  $r_c/r_{nc} \rightarrow 0$  (specifically, for  $r_c/r_{nc} = 0.1$ ,  $L_c(T_0)/\sigma_{c_c}^2 \approx 1.47$ , and  $L_{nc}(T_0)/\sigma_{c_{nc}}^2 \approx 1.21$ ), we conclude that the approximation is accurate even at fixed  $T^c = T_0$  when its assumptions are satisfied. This happens even though  $T_*^c$  depends on  $c_c$  and  $c_{nc}$ , but apparently the approximate estimates are as good as the ML estimates even at fixed  $T^c = T_0$  and work well for a large range of concentration ratios. This is important, as the molecular mechanisms that sets the delays in the cell does not need to be modified for different ligand concentrations.

In contrast, when the ligands are nearly indistinguishable ( $r_c/r_{nc} \sim 1$ ), both  $L_c(T_0)/\sigma_{c_c}^2 \sim 100$  and  $L_{nc}(T_0)/\sigma_{c_{nc}}^2 \sim 100$ , but here one would not use one receptor to estimate two concentrations since even the ML solution is bad (cf. Fig 2). Note also that both  $L_c$  and  $L_{nc}$  are smaller for  $r_c \sim r_{nc}$  if  $c_c \gg c_{nc}$ . This is because our main assumption (that almost all long events are cognate) holds better when cognate ligands dominate. Finally, the correlation coefficient between the approximate estimates,  $\rho^a$  (right panel) reaches -1 earlier than in Fig 2. This is a direct consequence of Eqs (13) and (14).

### Kinetic proofreading for approximate estimation

The approximate solution can be computed by cells using the well-known kinetic proofreading (KPR) mechanism [14, 15, 17, 18]. In the simplest model of KPR [19], intermediate states between an inactive and an active state of a receptor delay the activation. Thus bound ligands can dissociate before the receptor activates, at which point it quickly reverts to the inactive state. Since  $r_c < r_{nc}$ , cognate ligands dominate among bindings that persist to activation. The resulting increase in specificity in various KPR schemes has led to their exploration in the context of detection of rare ligands [11, 12, 16, 18]. Instead, here we analyze their ability to measure concentrations of both ligands simultaneously. We first consider the case where both the cognate and the non-cognate ligand concentration are comparable,  $c_c \sim c_{nc}$  and the dissociation rates are distinct,  $r_c \ll r_{nc}$ . In the following sections, we explore another case,  $c_c \ll c_{nc}$  and  $r_c \lesssim r_{nc}$ , a situation common in immunology.

Consider a biochemical network in Fig 4(a): the receptor, R, activates two messenger molecules, A and B. The former is activated with the rate  $k_A$  only if the receptor stays bound for longer than a certain  $T^c$  (with the delay achieved using the KPR intermediate states). The latter is activated with the rate  $k_B$  whenever the receptor is bound. The molecules deactivate with the rates  $r_A$  and  $r_B$ , respectively, and all activations/deactivations are first-order reactions. Then the mean concentrations of the messenger molecules are (see Methods):

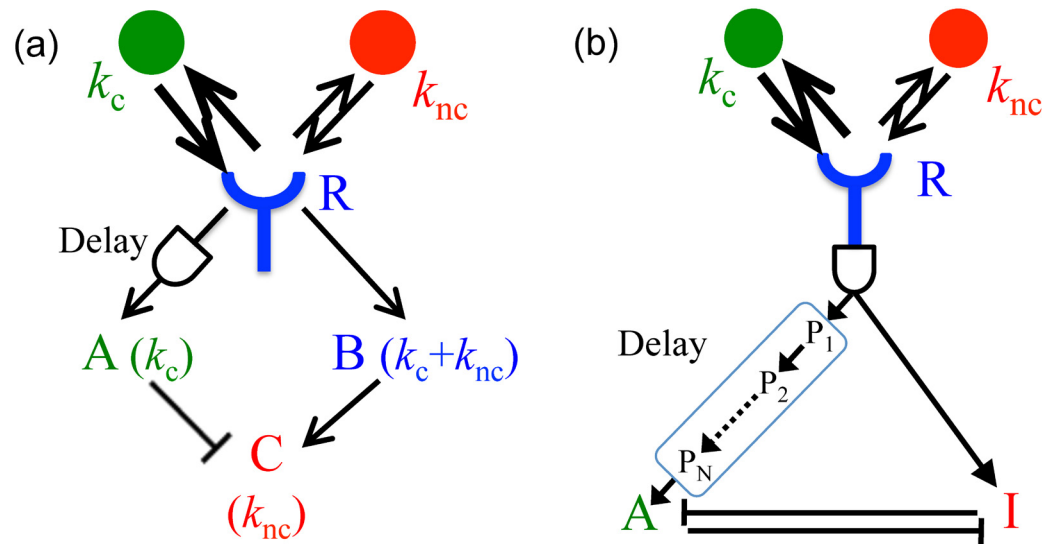
$$\bar{A} = \frac{k_c c_c / r_c e^{-r_c T^c} + k_{nc} c_{nc} / r_{nc} e^{-r_{nc} T^c}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} \frac{k_A}{r_A}, \tag{21}$$

$$\bar{B} = \frac{k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} \frac{k_B}{r_B}. \tag{22}$$

Assuming again that most bindings longer than  $T^c$  are cognate ( $T^c \gg 1/r_{nc}$ ), Eq (21), can be written as:

$$\bar{A} = \frac{k_c c_c / r_c e^{-r_c T^c}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} \frac{k_A}{r_A}. \tag{23}$$





**Fig 4. Kinetic proofreading for estimating multiple concentrations.** Two different kinetic schemes that allow using kinetic proofreading for estimation of chemical concentrations. (A) Molecules A and B are produced when the receptor R is bound, but A is produced only for long bindings, implemented through the KPR delay. Another chemical species C subtracts A from B, so that A approximates  $c_c$  and C approximates  $c_{nc}$ . This kinetic scheme would simultaneously estimate relatively similar concentrations of two ligands with substantially distinct off-rates  $r_{c,nc}$ . (B) Molecules A and I (inhibitor of A) are produced when the receptor is bound (possibly after a KPR delay). Production of A is delayed even further, as it results at the end of a sequence of intermediate products  $P_i$ . A and I bind to each other nearly irreversibly, sequestering and deactivating each other. The first delay at the receptor filters very short, nonspecific bindings. Because of the additional KPR delay on the A branch, A and I branches are then different linear combinations of strongly (cognate) and medium (non-cognate) binding ligands. If various kinetic parameters are tuned (see text), then even for very similar  $r_c$  and  $r_{nc}$ , and even for a rare cognate ligand, the sequestration can remove the non-cognate contribution from the A branch.

<https://doi.org/10.1371/journal.pcbi.1005490.g004>

Further, it is easy to see that Eq (22) can be rewritten as:

$$\frac{k_c c_c}{r_c} + \frac{k_{nc} c_{nc}}{r_{nc}} = \frac{\bar{B}}{k_B/r_B - \bar{B}}. \quad (24)$$

Now solving Eqs (23) and (24) for the on-rates, we get

$$c_c = \frac{\bar{A} e^{r_c T_c} r_c r_A}{k_c k_A} \left( 1 + \frac{\bar{B}}{k_B/r_B - \bar{B}} \right), \quad (25)$$

$$c_{nc} = \frac{r_{nc}}{k_{nc}} \left[ \frac{\bar{B}}{k_B/r_B - \bar{B}} - \frac{\bar{A} e^{r_c T_c} r_A}{k_A} \left( 1 + \frac{\bar{B}}{k_B/r_B - \bar{B}} \right) \right]. \quad (26)$$

The corrections of the form  $\bar{B}/(k_B/r_B - \bar{B})$  appear because bindings only happen to unbound receptors, as emphasized in Ref. [5]. However, these nonlinear relations are still hard to implement with simple biochemical components. We solve this by further assuming  $\epsilon = \bar{B}/(k_B/r_B) \ll 1$ , which is true if the receptor is mostly unbound, which happens at low

concentrations. This gives

$$c_c^{KPR} \approx \frac{\bar{A} e^{r_c T^c} r_c r_A}{k_c k_A}, \tag{27}$$

$$c_{nc}^{KPR} \approx \frac{r_{nc}}{k_{nc}} \left( \frac{r_B \bar{B}}{k_B} - \frac{\bar{A} e^{r_c T^c} r_A}{k_A} \right). \tag{28}$$

These equations are analogous to Eqs (13) and (14). They are easy to realize biochemically (cf. Fig 4(a)):  $c_c$  is related to the concentration of the proofread species A by a rescaling, and  $c_{nc}$  comes from subtracting rescaled versions of B and A from each other. The subtraction can be done by the third species C, activated by B and suppressed by A. Since  $\epsilon \ll 1$ , then  $\bar{A}$  and  $\bar{B}$  are small, and many such activation-suppression schemes are linearized as the subtraction [8]. Note that such incoherent feedforward loops (the receptor activates A and B, which then affect C incoherently by suppressing and activating it, respectively) are ubiquitous in cellular networks downstream of receptors [9].

The bias of  $c_c^a$  and  $c_{nc}^a$  due to long, but noncognate binding events, Eq (15), carries over to  $c_c^{KPR}$  and  $c_{nc}^{KPR}$ . However, there is an additional contribution since the time to traverse the intermediate states in KPR schemes with multiple intermediate steps is random. Thus  $T^c$  has some variance  $\sigma_{T^c}^2$  [19, 20]. This variability changes the rate of occurrence of long binding events, but they are still rare, nearly independent, and Poisson-distributed. Denoting by  $\langle \cdot \rangle$  the averaging at a fixed  $T^c$ , and by  $\bar{\cdot}$  the averaging over  $T^c$ , we get

$$\langle n_l \rangle = n \overline{P(\tau^b > T^c)} = \frac{n}{(k_c c_c + k_{nc} c_{nc})} \overline{(k_c c_c e^{-r_c T^c} + k_{nc} c_{nc} e^{-r_{nc} T^c})} \approx \frac{n k_c c_c e^{-r_c \bar{T}^c + \frac{1}{2} r_c^2 \sigma_{T^c}^2}}{(k_c c_c + k_{nc} c_{nc})}, \tag{29}$$

where we have used the approximation  $\bar{T}^c \gg 1/r_{nc}$  in the last step.

Thus  $\sigma_{T^c}^2$  effectively renormalizes the cutoff to  $\bar{T}^c - \frac{1}{2} r_c \sigma_{T^c}^2$ . Replacing  $T^c$  in Eqs (27) and (28) by its renormalized value, which is an easy change in the scaling factors, removes this additional bias due to the random  $T^c$  in the KPR scheme.

Since long bindings are rare, the variance of the KPR estimator is dominated again generally by  $\bar{A}$ , but not  $\bar{B}$ . The intrinsic stochasticity in the production of molecules of A contributes to the variance. However, this contribution can be made arbitrarily small by increasing  $k_A$ , and we neglect it here. A larger contribution comes from the random number of long bound intervals and a random duration of each of them. To calculate this, in the limit of rare long binding events, we use well-known results in the theory of noise propagation in chemical networks [21]

$$\frac{\sigma_A^2}{A^2} \approx \frac{(1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}) e^{r_c T^c - \frac{1}{2} r_c^2 \sigma_{T^c}^2}}{k_c c_c (1/r_c + 1/r_A)} = \frac{e^{r_c T^c - \frac{1}{2} r_c^2 \sigma_{T^c}^2}}{k_c c_c (1/r_c + 1/r_A)} + O(\epsilon). \tag{30}$$

This is a direct analog of Eq (16).

In principle, one can measure more than two concentrations similarly, as long as all species have distinct off-rates. For example, to estimate three concentrations, one needs an additional branch downstream of the receptor that proofreads for an intermediate time. Then the branches with the strongest, intermediate, and no proofreading would measure approximately the highest affinity ligand, a combination of the two higher affinity ligands, and all three ligands, respectively. Appropriate activation and inhibition of downstream targets will then allow identifying individual concentrations from these combined readouts. The error (the variance of the ML estimator, and both the bias and the variance for the approximate and the KPR estimators) would grow with an increasing number of ligand species, largely because a larger range of off-rates would be required to disambiguate more ligands. However, this would

still represent a dramatic increase in the information gained by a receptor that tracks its precise temporal dynamics, rather than just the average binding state.

### Using precise timing to disambiguate two similar ligands

Here we depart slightly from our scenario and show how a KPR-based scheme relying on the entire temporal sequence of activation / deactivation events can estimate the concentration of a *single* cognate ligand even if the two ligands have very similar off-rates  $r_c \lesssim r_{nc}$ , a situation common in immunology. In such a situation, the KPR branch gets activated not just by the cognate ligand, but also by the non-cognate ligand (though at a smaller rate). When the goal is the accurate estimation of the cognate ligand only, then the contribution to the KPR branch by the non-cognate ligand needs to be removed. To construct a signal transduction network able to do this, we abstract from the existing detailed model of FcεRI immunological receptor [9], a well studied eukaryotic signal transduction system mediating many allergic reactions [22]. Here the main signaling branch gets activated through the Lyn-Syk kinase pathway following kinetic proofreading after a ligand binds to the receptor [9]. However, receptor binding excites an additional branch early on, after only one step in kinetic proofreading (a single phosphorylation on the β chain of the receptor). This branch activates Inpp5d (SHIP) phosphatase, which later dephosphorylates Phosphatidylinositol 3-phosphate (PIP3), a key downstream output of the main signaling branch, and sequesters the dephosphorylated product PtdIns(3, 4)P<sub>2</sub> [9]. The part of this signaling motif relevant for our analysis is summarized in a deliberately simplified signaling diagram in Fig 4(b), where A stands for PtdIns(3, 4, 5)P<sub>3</sub> (PIP3), I stands for PtdIns(3, 4)P<sub>2</sub>, and I is produced by SHIP. Further, R is the FcεRI receptor bound to an antibody, and cognate and noncognate molecules are the antigens specific/nonspecific to the bound antibody.

In this network, we consider the main activator branch (A), activated after the usual KPR delay, and hence sensitive to long binding events only (which now have contributions both from  $k_c$  and  $k_{nc}$ ). The secondary inhibiting branch (I) is activated by many more binding events, though the shortest, nonspecific background binding events may be removed from both branches by additional proofreading steps (an early cross-phosphorylation event in the FcεRI system). The messengers in both branches later form a complex AI, and only A not in the complex activates further downstream signaling. If the production rates of A and I are appropriately matched (which can be done if the off-rates are known *a priori*, which they should be for such a molecular signal detection system), this sequestration of A by I can effectively remove the contribution to the A branch coming from the non-cognate ligand. The kinetic diagram can be described with the following rate equations (where, for simplicity, we neglect the first proofreading common to both branches):

$$\frac{dA}{dt} = \beta_A - r_A A - r_{AI} AI, \tag{31}$$

$$\frac{dI}{dt} = \beta_I - r_I I - r_{AI} AI, \tag{32}$$

where  $r_{A/I}$  are the degradation rates of the messengers A and I,  $r_{AI}$  is the sequestration rate, and  $\beta_{A/I}$  are the messenger production rates, derived as above:

$$\beta_A = \frac{k_c c_c / r_c e^{-r_c T^c} + k_{nc} c_{nc} / r_{nc} e^{-r_{nc} T^c}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} k_A, \tag{33}$$

$$\beta_I = \frac{k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} k_I. \tag{34}$$

Here  $k_{A/I}$  are the rates of production of A and I, respectively, when the receptor has been bound for a sufficiently long time to produce either.

We assume for simplicity  $r_A = r_I$ . Further, we choose  $r_A = r_I \ll r_{AI}A \sim r_{AI}I$ , so that sequestration rather than degradation is primarily responsible for the disappearance of the messengers. Then the steady state solution of the rate equations (Eqs 31 and 32) is [23] (see Methods):

$$\bar{A}_{ss} = \frac{(\beta_A - \beta_I)}{2r_I} + \sqrt{\left(\frac{\beta_A - \beta_I}{2r_I}\right)^2 + \frac{\beta_A}{r_{AI}}}, \tag{35}$$

$$\bar{I}_{ss} = \frac{(\beta_I - \beta_A)}{2r_I} + \sqrt{\left(\frac{\beta_I - \beta_A}{2r_I}\right)^2 + \frac{\beta_I}{r_{AI}}}. \tag{36}$$

The numerators of both  $\beta_A$  and  $\beta_I$  are linear combinations of  $c_c$  and  $c_{nc}$ . If the parameters of the biochemical networks are such that the production rate of the proofread branch is

$k_A = k_I e^{r_{nc} T^c}$ , then  $(\beta_A - \beta_I) = \frac{k_c c_c / r_c (e^{(r_{nc} - r_c) T^c} - 1)}{(1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc})} k_I > 0$ , which has a  $c_{nc}$ -independent numerator. Thus the contribution of non-cognate ligand to the activator branch is largely sequestered.

Moreover, for large  $r_{AI}$ , we have  $\bar{A}_{ss} \propto (1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc})^{-1}$ , so that the activation of the A branch *decreases* as  $c_{nc}$  increases. In contrast, if  $c_c = 0$  (no cognate ligands present), then

$\bar{A}_{ss} = \bar{I}_{ss} = \sqrt{\frac{k_I}{r_{AI}} \frac{k_{nc} c_{nc}}{k_{nc} c_{nc} + r_{nc}}}$ , which grows with  $c_{nc}$ . This behavior is reminiscent of the agonist-antagonist picture in FcεRI receptor activation [24]: a weak ligand by itself can activate the cellular response, but it inhibits (antagonizes) activation of the response by a stronger agonist if both are present.

## Discussion

The realization of Refs. [5, 13, 12] and others that the detailed temporal sequence of binding and unbinding events carries more information about the ligand concentration than the mean receptor occupancy is a conceptual breakthrough. It parallels the realization in the computational neuroscience community that precise timing of spikes carries more information about the stimulus than the mean neural firing rate [25–30], and it has a potential to be equally impactful. This extra information when measuring one ligand concentration with one receptor [5, 12] amounted to increasing the sensing accuracy by a constant prefactor, or, equivalently, getting only a finite number of additional bits from even a very long measurement [31]. In contrast, here we show that two concentrations can be measured with one receptor with the variance that decreases inversely proportionally to the number of observations,  $n$ , Eq (16), or to the integration time,  $1/r_B$ , Eq (30), so that the accuracy is only an (often small) prefactor lower than would be possible with one receptor per ligand species. Asymptotically, this doubles the information obtained by the receptor [31].

Crucially, such improvement would not be possible without the cross-talk, or binding among noncognate ligands and receptors. Normally, the cross-talk is considered a nuisance that must be suppressed [32, 33]. Instead, we argue that cross-talk can be beneficial by recruiting more receptor types to measure the concentration of the same ligand. In particular, this allows having fewer receptor than ligand species, potentially illuminating how cells function reliably in chemically complex environments with few receptor types. Further, the cross-talk can increase the dynamic range of the entire system: a ligand may saturate its cognate receptor,

preventing accurate measurement of its (high) concentration, but it may be in the sensitive range of non-cognate receptors at the same time. Finally, the increased bandwidth may lead to improvements in sensing a time-dependent ligand concentration [11, 13]. In forthcoming publications, we plan to explore such many-to-many sensory schemes, extending ideas of Ref. [34] to tracking temporal sequences of activation of the receptor and to temporally varying environments.

While the exact maximum likelihood inference of multiple concentrations from a temporal binding-unbinding sequence is rather complex, we showed that when the cognate and the non-cognate off-rates are substantially different, there is a simpler, approximate, but accurate inference procedure for joint measurements of cognate and noncognate ligands. And even if the off-rates are close, one can still measure the cognate ligand concentration reliably. Crucially, this inference can be performed by biochemical motifs readily available to the cell. Namely, one needs two branches of activation downstream of the receptor, with at least one of them having a kinetic proofreading (KPR) time delay. Then the individual ligand concentrations can be obtained by mutual inhibition between the two branches, or by incoherent feed-forward loops. We emphasize again that this allows estimation of *multiple* concentrations from activity of a single receptor, in contrast to a better estimation of just one concentration [12].

Our simple models only illustrate a wide class of models that can use the temporal structure of the receptor binding sequence to measure more than one ligand concentrations for various ligand combinations, including similar and dissimilar ligands. Additional branches from different points in the proofreading cascade provide additional information about the binding affinities of the mixture of ligands present in the environment, and then algebraic operations on these readouts can be performed by a large diversity of feedforward and feedback loops, competitions for the substrate and the enzyme, and so on. For example, in our simple model, the action of the antagonist is due to the competition for available receptors, while experiments suggest competition for a critical initiating kinase [24], which would require a straightforward modification of the model. Similarly, antagonists are usually “medium” affinity ligands, while very weak ligands do not antagonize receptors. As illustrated in Fig 4(b), this can be achieved by having an additional KPR time delay common to both A and I branches, which occurs in practice [9].

The kinetic diagram for the FcεRI receptor is not unique, and similar (though not equivalent) structures exist for other immune cells and receptors as well [9]. Such common structural features result in a similar phenomenology of activation profiles, which are different for pure ligands and ligand mixtures, and depend nontrivially on the details of the binding affinities and concentrations of the ligands in the mixture [10, 16, 35–38]. Interestingly, on longer time scales, a potentially related phenomenon in innate immune response is that of endotoxin tolerance (desensitization to commonly present ligands) [39], which also affects ligands of different affinity differently, and in this case also depends on the history of exposure to other ligands [40]. It is mediated by SHIP, a crucial player in our analysis of FcεRI signaling [41], whose activity may be interpreted as setting the relative gain on the A and I branches of Fig 4(b), thus resulting in a more accurate signal estimation. In other words, one interpretation of the known results is that, as various feedback loops increase the activity of SHIP in response to frequent activation of signaling downstream of the receptor, the amount of I increases, thus sequestering more A, lowering its steady-state activity, and inducing tolerance. An important contribution of the understanding developed here is that one can try to interpret these various kinetic diagrams and their phenomenological consequences as implementing estimation of concentrations of potentially many ligands (rather detection of a single one [11, 13, 16]), and maybe even doing it in a (nearly) Maximum Likelihood optimal fashion, under various assumptions about the number of distinct ligands, their relative abundance, and the (dis)

similarity of the off-rates. Exploring feasibility of such an interpretation is an additional interesting venue for future research.

In summary, monitoring precise temporal sequences of receptor activation/deactivation opens up new and exciting possibilities for environment sensing by cells.

## Methods

Here we provide mathematical derivations of some of the steps omitted in the *Results*.

### Derivation of maximum likelihood equations

We start with:

$$P \equiv P(\{\tau_i^b, \tau_i^u\} | c_c, c_{nc}) = \frac{1}{Z} \prod_{i=1}^n \left[ e^{-\tau_i^u(k_c c_c + k_{nc} c_{nc})} \left( k_c c_c r_c e^{-\tau_i^b r_c} + k_{nc} c_{nc} r_{nc} e^{-\tau_i^b r_{nc}} \right) \right]. \quad (37)$$

The log-likelihood of  $k_{c,nc}$  is the logarithm of  $P$ :

$$\log(P) = -\log Z - \sum_{i=1}^n \left[ \tau_i^u (k_c c_c + k_{nc} c_{nc}) + \log \left( k_c c_c r_c e^{-\tau_i^b r_c} + k_{nc} c_{nc} r_{nc} e^{-\tau_i^b r_{nc}} \right) \right]. \quad (38)$$

Taking the derivatives of the log-likelihood w. r. t.  $c_c$  and  $c_{nc}$  and setting them to zero gives the Maximum Likelihood (ML) equations for the concentrations. These are:

$$\frac{\partial \log(P)}{\partial c_c} = - \sum_{i=1}^n \tau_i^u k_c + \sum_{i=1}^n \frac{k_c r_c e^{-\tau_i^b r_c}}{D(c_c^*, c_{nc}^*, \tau_i^b)} = 0, \quad (39)$$

$$\frac{\partial \log(P)}{\partial c_{nc}} = - \sum_{i=1}^n \tau_i^u k_{nc} + \sum_{i=1}^n \frac{k_{nc} r_{nc} e^{-\tau_i^b r_{nc}}}{D(c_c^*, c_{nc}^*, \tau_i^b)} = 0. \quad (40)$$

Here,  $D(c_c^*, c_{nc}^*, \tau_i^b) = (k_c c_c^* r_c e^{-\tau_i^b r_c} + k_{nc} c_{nc}^* r_{nc} e^{-\tau_i^b r_{nc}})$ , with \* denoting the ML solution.

Denoting by  $T^u = \sum_{i=1}^n \tau_i^u$ , the total time for which the receptor is unbound, these equations can be rewritten as

$$-k_c T^u + \sum_{i=1}^n \frac{k_c r_c e^{-\tau_i^b r_c}}{D(k_c^*, k_{nc}^*, \tau_i^b)} = 0, \quad (41)$$

$$-k_{nc} T^u + \sum_{i=1}^n \frac{k_{nc} r_{nc} e^{-\tau_i^b r_{nc}}}{D(k_c^*, k_{nc}^*, \tau_i^b)} = 0. \quad (42)$$

Multiplying Eqs (41) and (42) by  $c_c^*$  and  $c_{nc}^*$ , respectively, and adding them gives

$$k_c c_c^* + k_{nc} c_{nc}^* = \frac{n}{T^u}. \quad (43)$$

### Comparison of simulations with analytical results for single concentration estimation

Here we compare the results obtained from the numerical simulations to the analytical expressions derived in Ref. [12] for detection of the concentration of the cognate ligand in a background of spurious ligands. The variance of the concentration estimation obtained from the

simulations matches quite well with integral expression, Eq. (7) in Ref. [12], Fig 5a. Note that this expression is inverse of the (1, 1) term of the Hessian matrix, Eq 7. The analytical results obtained for the low concentration of the cognate ligand compared to a non cognate ligand ( $c_c \ll c_{nc}$ ) also match the simulations, Fig 5b.

### Approximate solution

**Derivation of Eqs 13 and 14.** Defining long events as  $\tau_i^b \geq T^c$  and using Eq (43), we rewrite Eq (41) as

$$\frac{k_c n}{k_c c_c^* + k_{nc} c_{nc}^*} = \left( \sum_{\tau_i^b \geq T^c} + \sum_{\tau_i^b < T^c} \right) \frac{k_c r_c e^{-\tau_i^b r_c}}{\left( k_c c_c^* r_c e^{-\tau_i^b r_c} + k_{nc} c_{nc}^* r_{nc} e^{-\tau_i^b r_{nc}} \right)}. \tag{44}$$

Assuming that all long events are cognate,  $T^c \gg 1/r_{nc}$ , we can ignore the  $k_{nc} c_{nc}^* r_{nc} e^{-\tau_i^b r_{nc}}$  in the denominator in the first sum. This gives

$$\frac{k_c n}{k_c c_c^a + k_{nc} c_{nc}^a} = \frac{n_1}{c_c^a} + \sum_{\tau_i^b < T^c} \frac{k_c r_c e^{-\tau_i^b r_c}}{D(c_c^a, c_{nc}^a, \tau_i^b)},$$

where  $n_1$  is the number of long events, and the superscript ‘‘a’’ stands for the approximate solution. If further  $T$  is long enough so that there are many short events, and a single binding duration hardly affects  $k_c^*$ , then the sum in Eq (9) can be approximated by the expectation value:

$$\frac{n}{k_c c_c^a + k_{nc} c_{nc}^a} = \frac{n_1}{k_c c_c^a} + (n - n_1) \int_0^{T^c} \frac{r_c e^{-\tau^b r_c} P(\tau^b | c_c^a, c_{nc}^a) d\tau^b}{D(c_c^a, c_{nc}^a, \tau^b)}, \tag{45}$$

where  $P(\tau^b | c_c^a, c_{nc}^a)$  is the probability of observing a binding event of duration  $\tau^b$  for the given binding rates,

$$P(\tau^b | c_c^a, c_{nc}^a) = \frac{D(c_c^a, c_{nc}^a, \tau^b)}{(k_c c_c^a + k_{nc} c_{nc}^a)}. \tag{46}$$

Plugging Eq (11) into Eq (10), we obtain

$$\frac{n}{k_c c_c^a + k_{nc} c_{nc}^a} = \frac{n_1}{k_c c_c^a} + (n - n_1) \int_0^{T^c} \frac{r_c e^{-\tau^b r_c} d\tau^b}{(k_c c_c^a + k_{nc} c_{nc}^a)}, \tag{47}$$

which gives:

$$\frac{1}{(k_c c_c^a + k_{nc} c_{nc}^a)} = \frac{n_1}{n k_c c_c^a} + \left( 1 - \frac{n_1}{n} \right) \frac{1 - e^{-r_c T^c}}{(k_c c_c^a + k_{nc} c_{nc}^a)}. \tag{48}$$

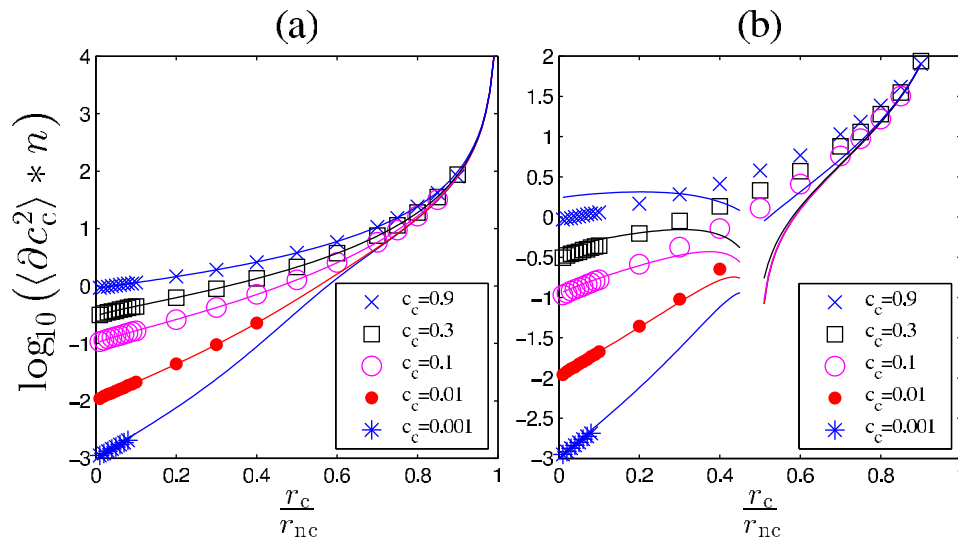
Assuming  $n_1 \ll n$ , we get:

$$\frac{1}{(k_c c_c^a + k_{nc} c_{nc}^a)} = \frac{n_1}{n k_c c_c^a} + \frac{1}{(k_c c_c^a + k_{nc} c_{nc}^a)} - \frac{e^{-r_c T^c}}{(k_c c_c^a + k_{nc} c_{nc}^a)}. \tag{49}$$

This gives,

$$c_c^a = \frac{1}{k_c} \frac{n_1}{T^c} e^{r_c T^c}. \tag{50}$$





**Fig 5. Comparison of simulations with analytical results for single concentration estimation.** (a) The variance of the estimated concentration in simulations (markers) matches quite well with the integral expression, Eq. (7) in Ref. [12]. (b) Comparison of simulations to analytical expression derived for low concentrations in Eq. (9) of Ref. [12].

<https://doi.org/10.1371/journal.pcbi.1005490.g005>

Finally, using Eq 43 we get

$$c_{nc}^a = \frac{1}{k_{nc}} \left( \frac{n}{T^u} - \frac{n_1}{T^u} e^{r_c T^c} \right). \tag{51}$$

**Error in approximate solution at  $T^c$ .** Plugging Eq 20 in Eq 15 we get,

$$\begin{aligned} \langle c_c^a \rangle &\approx c_c + \frac{k_{nc} c_{nc}}{k_c} e^{-\frac{(r_{nc}-r_c)}{(2r_{nc}-r_c)}} \log \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right] \\ &= c_c + \frac{k_{nc} c_{nc}}{k_c} \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right]^{\frac{(r_{nc}-r_c)}{(2r_{nc}-r_c)}} \end{aligned} \tag{52}$$

In the limit  $r_{nc} \gg r_c$ , the second term goes as  $\sqrt{c_c}$ .

Now using this in Eq 16, we get:

$$\begin{aligned} \sigma^2(c_c^a) &\approx \langle c_c^a \rangle \frac{(c_c + k_{nc} c_{nc} / k_c)}{n} e^{\frac{r_c}{(2r_{nc}-r_c)}} \log \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right] \\ &= \left( c_c + \frac{k_{nc} c_{nc}}{k_c} \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right]^{\frac{(r_{nc}-r_c)}{(2r_{nc}-r_c)}} \right) \frac{(c_c + k_{nc} c_{nc} / k_c)}{n} \left[ 2T^u \left( \frac{r_{nc}}{r_c} - 1 \right) \frac{k_{nc}^2 c_{nc}^2}{k_c c_c} \right]^{\frac{r_c}{(2r_{nc}-r_c)}}. \end{aligned} \tag{53}$$

**Kinetic proofreading for approximate estimation: Derivation of Eqs (21) and (22)**

In the biochemical network in Fig 4(a) of the main text, the receptor R activates two messenger molecules, A and B. The former is activated with the rate  $k_A$  only if the receptor stays bound

for longer than a certain  $T^c$  (with the delay achieved using the KPR intermediate states). The latter is activated with the rate  $k_B$  whenever the receptor is bound. The molecules deactivate with the rates  $r_A$  and  $r_B$ , respectively, and all activations/deactivations are first-order reactions. The rate equation for the two molecules can be written as:

$$\frac{dA}{dt} = k_A \Theta(\tau_b > T^c) - r_A A, \quad (54)$$

$$\frac{dB}{dt} = k_B \Theta(\tau_b > 0) - r_B B. \quad (55)$$

The  $\Theta$  functions represent the fact that  $A$  is produced only when the receptor has been bound for longer than the cutoff time  $T^c$ , and  $B$  is produced only when the receptor is bound.

The steady state value of  $\bar{A}$  can be obtained by equating the average deactivation rate  $r_A \bar{A}$  to  $k_A$  times the fraction of time the receptor occupancy was larger than the cutoff,  $T^c$ , i.e.,

$$r_A \bar{A} = k_A \frac{\langle \tau_b > T^c \rangle}{\langle \tau_u \rangle + \langle \tau_b \rangle}. \quad (56)$$

Similarly,  $\bar{B}$  can be obtained as:

$$r_B \bar{B} = k_B \frac{\langle \tau_b \rangle}{\langle \tau_u \rangle + \langle \tau_b \rangle}. \quad (57)$$

Therefore, the mean concentrations of the messenger molecules are:

$$\bar{A} = \frac{k_c c_c / r_c e^{-r_c T^c} + k_{nc} c_{nc} / r_{nc} e^{-r_{nc} T^c} k_A}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} \frac{k_A}{r_A}, \quad (58)$$

$$\bar{B} = \frac{k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}}{1 + k_c c_c / r_c + k_{nc} c_{nc} / r_{nc}} \frac{k_B}{r_B}. \quad (59)$$

### Using precise timing to disambiguate two close ligands: Derivation of Eqs (35) and (36)

The rate equations are:

$$\frac{dA}{dt} = \beta_A - r_A A - r_{AI} AI, \quad (60)$$

$$\frac{dI}{dt} = \beta_I - r_I I - r_{AI} AI. \quad (61)$$

Equating the r. h. s. to zero gives the steady state conditions:

$$\beta_A - r_A \bar{A}_{ss} - r_{AI} \bar{A}_{ss} \bar{I}_{ss} = 0, \quad (62)$$

$$\beta_I - r_I \bar{I}_{ss} - r_{AI} \bar{A}_{ss} \bar{I}_{ss} = 0. \quad (63)$$

The latter of these can be rewritten as:

$$\bar{I}_{ss} = \frac{\beta_I}{r_I + r_{AI}\bar{A}_{ss}}. \quad (64)$$

Plugging this in Eq (62), we get

$$\beta_A - r_A\bar{A}_{ss} - \frac{r_{AI}\bar{A}_{ss}\beta_I}{r_I + r_{AI}\bar{A}_{ss}} = 0, \quad (65)$$

which can be simplified to:

$$\bar{A}_{ss}^2 + \left(\frac{r_I}{r_{AI}} + \frac{(\beta_I - \beta_A)}{r_A}\right)\bar{A}_{ss} - \frac{\beta_A r_I}{r_{AI}r_A} = 0. \quad (66)$$

This quadratic equation has the solution:

$$\bar{A}_{ss} = -\left(\frac{r_I}{2r_{AI}} + \frac{(\beta_I - \beta_A)}{2r_A}\right) + \sqrt{\left(\frac{r_I}{2r_{AI}} + \frac{(\beta_I - \beta_A)}{2r_A}\right)^2 + \frac{\beta_A r_I}{r_{AI}r_A}}. \quad (67)$$

Now assuming  $r_A = r_I$  and  $r_A = r_I \ll r_{AI}A \sim r_{AI}I$ , we get:

$$\bar{A}_{ss} = \frac{(\beta_A - \beta_I)}{2r_I} + \sqrt{\left(\frac{\beta_I - \beta_A}{2r_A}\right)^2 + \frac{\beta_A}{r_{AI}}} \quad (68)$$

One can similarly get the equation for  $\bar{I}_{ss}$  as well.

## Acknowledgments

We thank Rustom Antia, Lily Chylek, William Hlavacek, Andre Levchenko, Andrew Mugler, Dmitry Shayakhmetov, and Veronika Zarnitsyna for useful discussions.

## Author Contributions

**Conceptualization:** VS IN.

**Formal analysis:** VS IN.

**Funding acquisition:** IN.

**Investigation:** VS IN.

**Methodology:** VS IN.

**Project administration:** IN.

**Software:** VS IN.

**Supervision:** IN.

**Validation:** VS IN.

**Visualization:** VS.

**Writing – original draft:** VS IN.

**Writing – review & editing:** VS IN.

## References

1. Berg H, Purcell E. Physics of chemoreception. *Biophys J*. 1977; 20:193–219. [https://doi.org/10.1016/S0006-3495\(77\)85544-6](https://doi.org/10.1016/S0006-3495(77)85544-6) PMID: 911982
2. Bray D, Levin MD, Morton-Firth CJ. Receptor clustering as a cellular mechanism to control sensitivity. *Nature*. 1998; 393(6680):85–88. <https://doi.org/10.1038/30018> PMID: 9590695
3. Bialek W, Setayeshgar S. Physical limits to biochemical signaling. *Proc Natl Acad Sci (USA)*. 2005; 102:10040–10045. <https://doi.org/10.1073/pnas.0504321102>
4. Endres R, Wingreen N. Accuracy of direct gradient sensing by single cells. *Proc Natl Acad Sci (USA)*. 2008; 105:15749–15754. <https://doi.org/10.1073/pnas.0804688105> PMID: 18843108
5. Endres RG, Wingreen NS. Maximum likelihood and the single receptor. *Phys Rev Lett*. 2009; 103(15):158101. <https://doi.org/10.1103/PhysRevLett.103.158101> PMID: 19905667
6. Hu B, Chen W, Rappel WJ, Levine H. Physical limits on cellular sensing of spatial gradients. *Phys Rev Lett*. 2010; 105(4):048104. <https://doi.org/10.1103/PhysRevLett.105.048104> PMID: 20867888
7. Kaizu K, de Ronde W, Pajjmans J, Takahashi K, Tostevin F, ten Wolde PR. The berg-purcell limit revisited. *Biophys J*. 2014; 106(4):976–985. <https://doi.org/10.1016/j.bpj.2013.12.030> PMID: 24560000
8. Ellison D, Mugler A, Brennan M, Lee S, Huebner R, Shamir E, et al. Cell-cell communication can enhance the effect of shallow gradients of cues guiding cell growth and morphogenesis. *Proc Natl Acad Sci (USA)*. 2016; 113:E679–E688. <https://doi.org/10.1073/pnas.1516503113> PMID: 26792522
9. Chylek LA, Holowka DA, Baird BA, Hlavacek WS. An interaction library for the FcεRI signaling network. *Frontiers in Immunology*. 2014; 5. <https://doi.org/10.3389/fimmu.2014.00172> PMID: 24782869
10. Hlavacek WS, Redondo A, Wofsy C, Goldstein B. Kinetic proofreading in receptor-mediated transduction of cellular signals: receptor aggregation, partially activated receptors, and cytosolic messengers. *Bulletin of mathematical biology*. 2002 Sep; 64(5):887–911. <https://doi.org/10.1006/bulm.2002.0306> PMID: 12391861
11. Lalanne JB, Francois P. Chemodetection in fluctuating environments: receptor coupling, buffering, and antagonism. *Proc Natl Acad Sci (USA)*. 2015; 112(6):1898–1903. <https://doi.org/10.1073/pnas.1420903112> PMID: 25624502
12. Mora T. Physical Limit to Concentration Sensing Amid Spurious Ligands. *Physical Review Letters*. 2015 Jul; 115(3):038102. <https://doi.org/10.1103/PhysRevLett.115.038102> PMID: 26230828
13. Siggia ED, Vergassola M. Decisions on the fly in cellular sensory systems. *Proceedings of the National Academy of Sciences*. 2013; 110(39):E3704–E3712. <https://doi.org/10.1073/pnas.1314081110> PMID: 24019464
14. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci (USA)*. 1974; 71(10):4135–4139. <https://doi.org/10.1073/pnas.71.10.4135> PMID: 4530290
15. Ninio J. Kinetic amplification of enzyme discrimination. *Biochimie*. 1975; 57(5):587–595. [https://doi.org/10.1016/S0300-9084\(75\)80139-8](https://doi.org/10.1016/S0300-9084(75)80139-8) PMID: 1182215
16. Francois P, Voisinne G, Siggia E, Altan-Bonnet G, Vergassola M. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *Proc Natl Acad Sci (USA)*. 2013; 110(10):E888–97. <https://doi.org/10.1073/pnas.1300752110> PMID: 23431198
17. McKeithan TW. Kinetic proofreading in T-cell receptor signal transduction. *Proc Natl Acad Sci (USA)*. 1995; 92(11):5042–5046. <https://doi.org/10.1073/pnas.92.11.5042> PMID: 7761445
18. Goldstein B, Coombs D, Faeder JR, Hlavacek WS. Kinetic proofreading model. *Adv Exp Med Biol*. 2008; 640:82–94. [https://doi.org/10.1007/978-0-387-09789-3\\_8](https://doi.org/10.1007/978-0-387-09789-3_8) PMID: 19065786
19. Bel G, Munsky B, Nemenman I. The simplicity of completion time distributions for common complex biochemical processes. *Phys Biol*. 2010; 7:0610003. <https://doi.org/10.1088/1478-3975/7/1/016003> PMID: 20026876
20. Cheng X, Merchan L, Tchernookov M, Nemenman I. A large number of receptors may reduce cellular response time variation. *Phys Biol*. 2013; 10(3):035008. <https://doi.org/10.1088/1478-3975/10/3/035008> PMID: 23735700
21. Paulsson J. Models of stochastic gene expression. *Phys Life Rev*. 2005;p. 157–175. <https://doi.org/10.1016/j.plev.2005.03.003>
22. Faeder JR, Hlavacek WS, Reischl I, Blinov ML, Metzger H, Redondo A, et al. Investigation of early events in Fc epsilon RI-mediated signaling using a detailed mathematical model. *Journal of immunology (Baltimore, Md: 1950)*. 2003 Apr; 170(7):3769–3781. PMID: 12646643
23. Sprinzak D, Lakhanpal A, LeBon L, Santat LA, Fontes ME, Anderson GA, et al. Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature*. 2010; 465(7294):86–90. <https://doi.org/10.1038/nature08959> PMID: 20418862

24. Torigoe C, Inman JK, Metzger H. An unusual mechanism for ligand antagonism. *Science* (New York, NY). 1998 Jul; 281(5376):568–572. <https://doi.org/10.1126/science.281.5376.568>
25. Strong S, Koberle R, Steveninck RRdRv, Bialek W. Entropy and information in neural spike trains. *Phys Rev Lett*. 1998; 80(1):197–200. <https://doi.org/10.1103/PhysRevLett.80.197>
26. Reinagel P, Reid RC. Temporal coding of visual information in the thalamus. *J Neurosci*. 2000; 20(14):5392–5400. PMID: [10884324](https://pubmed.ncbi.nlm.nih.gov/10884324/)
27. Liu R, Tzonev S, Rebrik S, Miller K. Variability and information in a neural code of the cat lateral geniculate nucleus. *J Neurophysiol*. 2001; 86:2789–2806. PMID: [11731537](https://pubmed.ncbi.nlm.nih.gov/11731537/)
28. Nemenman I, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Comp Biol*. 2008; 4(3):e1000025. <https://doi.org/10.1371/journal.pcbi.1000025> PMID: [18369423](https://pubmed.ncbi.nlm.nih.gov/18369423/)
29. Fairhall A, Shea-Brown E, Barreiro A. Information theoretic approaches to understanding circuit function. *Curr Opin Neurobiol*. 2012; 22:653–659. <https://doi.org/10.1016/j.conb.2012.06.005> PMID: [22795220](https://pubmed.ncbi.nlm.nih.gov/22795220/)
30. Tang C, Chehayeb D, Srivastava K, Nemenman I, Sober SJ. Millisecond-scale motor encoding in a cortical vocal area. *PLoS Biol*. 2014; 12(12):e1002018. <https://doi.org/10.1371/journal.pbio.1002018> PMID: [25490022](https://pubmed.ncbi.nlm.nih.gov/25490022/)
31. Bialek W, Nemenman I, Tishby N. Predictability, complexity, and learning. *Neural Comput*. 2001; 13:2409. <https://doi.org/10.1162/089976601753195969> PMID: [11674845](https://pubmed.ncbi.nlm.nih.gov/11674845/)
32. McClean M, Mody A, Broach J, Ramanathan S. Cross-talk and decision making in MAP kinase pathways. *Nat Gen*. 2007; 39(3):409–414. <https://doi.org/10.1038/ng1957> PMID: [17259986](https://pubmed.ncbi.nlm.nih.gov/17259986/)
33. Laub M, Goulian M. Specificity in Two-Component Signal Transduction Pathways. *Ann Rev Genet*. 2007; 41:121–145. <https://doi.org/10.1146/annurev.genet.41.042007.170548> PMID: [18076326](https://pubmed.ncbi.nlm.nih.gov/18076326/)
34. Tsiron J, Ault A, Broach J, Morozov A. Decoding complex chemical mixtures with a physical model of a sensor array. *PLoS Comput Biol*. 2011; 7:e1002224. <https://doi.org/10.1371/journal.pcbi.1002224> PMID: [22046111](https://pubmed.ncbi.nlm.nih.gov/22046111/)
35. Cowley D, Schulze A. Conformational dynamics and kinetics of peptide antagonist interactions with interleukin-1 receptor. Fluorescence studies using the NBD-labelled peptide AF12415. *J Pept Res*. 1997; 49:444–454. <https://doi.org/10.1111/j.1399-3011.1997.tb00897.x> PMID: [9211226](https://pubmed.ncbi.nlm.nih.gov/9211226/)
36. Liu Z, Haleem-Smith H, Chen H, Metzger H. Unexpected signals in a system subject to kinetic proof-reading. *Proc Natl Acad Sci (USA)*. 2001; 98:7289–7294. <https://doi.org/10.1073/pnas.121171998> PMID: [11371625](https://pubmed.ncbi.nlm.nih.gov/11371625/)
37. Stefanova I, Hemmer B, Vergelli M, Martin R, Biddison W, Germain R. TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways. *Nat Immunol*. 2003; 4:248–254. <https://doi.org/10.1038/ni895> PMID: [12577055](https://pubmed.ncbi.nlm.nih.gov/12577055/)
38. Wylie D, Das J, Chakraborty A. Sensitivity of T cells to antigen and antagonism emerges from differential regulation of the same molecular signaling module. *Proc Natl Acad Sci (USA)*. 2007; 104:5533–5538. <https://doi.org/10.1073/pnas.0611482104> PMID: [17360359](https://pubmed.ncbi.nlm.nih.gov/17360359/)
39. Medzhitov R, Schneider D, Soares M. Disease Tolerance as a Defense Strategy. *Science*. 2012; 335:936–941. <https://doi.org/10.1126/science.1214935> PMID: [22363001](https://pubmed.ncbi.nlm.nih.gov/22363001/)
40. Chen K, Geng S, Yuan R, Diao N, Upchurch Z, Li L. Super-low dose endotoxin pre-conditioning exacerbates sepsis mortality. *EBioMedicine*. 2015; 2:324–333. <https://doi.org/10.1016/j.ebiom.2015.03.001> PMID: [26029736](https://pubmed.ncbi.nlm.nih.gov/26029736/)
41. Sly L, Rauh M, Kalesnikoff J, Song C, Krystal G. LPS-Induced Upregulation of SHIP Is Essential for Endotoxin Tolerance. *Immunity*. 2004; 21:227–239. <https://doi.org/10.1016/j.immuni.2004.07.010> PMID: [15308103](https://pubmed.ncbi.nlm.nih.gov/15308103/)