Information theory of adaptation in neurons, behavior, and mood

Tatyana O Sharpee[1], Adam J Calhoun[1,2] and Sreekanth H Chalasani[2]

The ability to make accurate predictions of future stimuli and consequences of one's actions are crucial for the survival and appropriate decision-making. These predictions are constantly being made at different levels of the nervous system. This is evidenced by adaptation to stimulus parameters in sensory coding, and in learning of an up-to-date model of the environment at the behavioral level. This review will discuss recent findings that actions of neurons and animals are selected based on detailed stimulus history in such a way as to maximize information for achieving the task at hand. Information maximization dictates not only how sensory coding should adapt to various statistical aspects of stimuli, but also that reward function should adapt to match the predictive information from past to future.

Addresses
[1] Computational Neurobiology Laboratory, Salk Institute for Biological Studies, United States
[2] Molecular Neurobiology Laboratory, Salk Institute for Biological Studies, United States

Corresponding author: Sharpee, Tatyana O (sharpee@salk.edu)

Recently Stephen Hawking cautioned against efforts to contact aliens [1], such as by beaming songs into space, saying: "We only have to look at ourselves to see how intelligent life might develop into something we wouldn't want to meet." Although one might wonder why we should ascribe the characteristics of human behavior to aliens, it is plausible that the rules of behavior are not arbitrary but might be general enough to not depend on the underlying biological substrate. Specifically, recent theories posit that the rules of behavior should follow the same fundamental principle of acquiring information about the state of environment in order to make the best decisions based on partial data [2••,3]. Further, these principles could also incorporate both the cost of obtaining information and the cost of making complex decisions [4••]. Therefore, validating such theories could help establish frameworks to compare behavior not only in different species and tasks, but also in single cells [5], neurons, intracellular pathways, as well as emergent phenomena at the population level, such as the distribution of blood flow in the brain that anticipates future stimuli [6•] as well as resource allocation within companies and government [7].

In this article, we review recent evidence that behavior in different systems can be described within a common framework whereby actions are chosen to maximize the Shannon mutual information with respect to a variable that quantifies performance in the task at hand. This idea has a venerable history when applied to individual neurons. In this case the mutual information represents how well the neural responses encode incoming stimuli, reviewed in [2••]. The mutual information can be computed as the difference between the entropy of the neural response $H(r)$ and the average entropy $\langle H(r|stim) \rangle$ of the neural response observed when a given stimulus is repeated multiple times, averaged over different stimuli:

$$I = H(r) - \langle H(r|stim) \rangle_{stim}$$

Adaptive changes in neural representation can be viewed as a predictive computation about the properties of stimuli to be received in the near future. By properly allocating the inherently limited neural responses through mechanisms such as adjusting the neural gain in single neurons [8,9] or the distribution of the preferred stimulus values for different neurons [10–12], neurons can more accurately encode future stimuli in order to provide more information about them. Indeed, adaptation has been shown to directly increase [13] and maintain [14] information transmission.

The gain of single neuron responses can adapt to increasingly subtle statistical properties, from the mean and variance of incoming stimuli (review of the earlier work is provided in [8,15]), to the detailed structure of sound sequences that extend over minutes [16] or to a power spectrum [17,18] and facial features in the visual system [19,20]. Similarly, maximally informative encoding differs strikingly depending upon higher-order stimulus statistics [21]. For Gaussian inputs, the classic perceptron-like solutions where the neural response depends only on one stimulus dimension provide maximum information. In contrast, in the case of Laplacian inputs, which approximate many of the inputs derived from the natural sensory environment [2••], the maximally informative solution prescribes that responses of single neurons should depend in specific and nonlinear ways on multiple image components. The corresponding nonlinearities were strikingly similar to those observed experimentally in the primary visual cortex with respect to three cone

isolating inputs [22•] as well as the relevant gray scale features [23]. At the level of neural populations, theoretical studies have described how the neural responses and their variability can be coordinated between neurons in such a way as to maximize information transmission [11,12,24,25]. For instance, either positive or negative noise correlations can improve the separability the neural responses to different classes of inputs, if the mean neural responses are negatively or positive correlated across different inputs [26]. These predictions were recently confirmed in experiments addressing changes in neuronal correlations through adaptation [27,28], learning [29] or specific behavioral paradigms, such as parturition [30]. While the neural mechanisms underlying coordination between neurons both in terms of their mean responses and their variability remain to be fully elucidated, the formation of synaptic connections is likely to play an important role. To this end, dendrites of single neurons in the developing cortex were shown to be capable of complex adaptive computations to navigate in shallow chemical gradients to ensure optimal connectivity [31,32]. In the retina, the resulting mosaics of receptive fields is such that even irregularities in the light sensitive regions of retinal ganglion cells carry substantial information [10,33–35].

Yet, the pervasive nature of adaptation and optimality observed at a variety of levels in the nervous system should not be taken for granted. Some statistical parameters, such as kurtosis, do not seem to trigger adaptive changes in neuronal gain [36••,37]. They however can affect the time scale of adaptation to other parameters of the stimulus distribution, such as variance [36••]. Furthermore, the ability of neurons to adapt to changes in the variance is not present in newborn neurons and develops slowly over the course of several weeks [38]. This staggering of different adaptive capabilities during development points to the computational cost associated with adaptation itself. It seems possible to quantify this cost in information-theoretic terms, perhaps similarly to how complexity of decisions and rewards was quantified [4••]. This would then make it possible to analyze the trade-off between creating a more fully adapted neural representation versus the gain in information transmission that this representation might provide. Such a framework may help explain when it is better to use automatic forms of adaptation, such as those that are due to built-in nonlinearities in the system [39,40] versus more flexible forms of adaptation where the adaptation time itself reflects the process of statistical inference [36••]. Ultimately, this may provide an explanation for adaptive properties in more natural stimulus ensembles [41•,42•,43•].

It turns out that that information maximization can also accurately describe changes in decision making observed at the behavioral level. Our first example in this area pertains to navigation in a turbulent environment. We know that tiny moths can find their mates over the distances of several kilometers based on encountering just a few pheromone molecules [44]. A computational algorithm for successful navigation in a turbulent environment was recently found [45••] by choosing trajectories that maximize the amount of information about the source location, although individual steps do not always bring the animal closer to the target. These 'infotaxis' trajectories reproduced characteristic properties of moth flight, such as cross-wind zigzagging far from the source and following the increasing odorant gradient close to the source. Note that the distinguishing feature of infotaxis is that, at low signal-to-noise conditions, its operation does not rely on time averaged stimulus characteristics, such as mean gradient. This property has its counterparts in neural adaptation, where for example, at low light levels, the gain of retinal ganglion cells can be affected by absorption of single photons [46•], a point to which we will return later.

From a physical perspective, where humans choose to look is a very different type of behavior than moth navigation, yet similar principles appear to be at work. Recent work has explained many aspects of human eye movements [47••], including looking in between two likely locations ('center-of-gravity saccades' [48]) or looking away from the target to eliminate less likely target locations, with an optimization strategy termed 'ideal observer analysis'. It turns out that the ideal observer analysis in this context can be mapped, in a one-to-one fashion, onto the problem of maximizing the acquisition of information about the target location. Thus, despite different physical constraints and behavioral goals, both the search with eye movements for a visual target and moth navigation using pheromones can be understood as information maximizing search strategies. It is noteworthy that the statistics of our eye movements is also under adaptive control to satisfy the needs of a trade-off between speed and accuracy (such as in reading) [49].

Encouraged by this success of the infotaxis framework [45,47••], we explored whether it could also account for yet another very different type of animal behavior, namely how a small nematode worm *C. elegans* decides to stop searching a local area for food. Searching for food over areas much larger in scale than the body size is a problem that many different types of species have to solve. A key feature of the infotaxis strategy is that information is continuously gained from both the presence and absence of odorant detection events. The goal is to maximize the function [45••]:

$$\Delta S(r_{new}|r_{current}) = -P(r_{current})S_{current} + (1 - P)$$
$$\times \; r_{current} \sum_{n=0}^{\alpha} p_n(r_{new})\Delta S_n(r_{new}),$$

where $P(r)$ describes the probability to find a source at location r, the entropy of this distribution is denoted as $S_{current}$, and the current position of the searcher as $r_{current}$. The terms p_n describe the expected probability to observe n odorant hits if the searcher decides to move to a location r_{new}, whereas terms ΔS_n describe the corresponding expected change in entropy following these outcomes. By comparison, a chemotaxis search would instead maximize the mean number of expected odorant detection events: $\sum_{n=0}^{\alpha} n\, p_n$. Another important distinction between infotaxis with respect to chemotaxis lies in the computation of $p_n(r_{new})$ In the infotaxis model this probability is updated following each odorant detection events and depends of the times where these events have occurred, whereas for the chemotaxis computation this probability only depends on the current position. At high odorant concentrations, such as those that often occur close to the source, infotaxis converges to simple chemotaxis. The key difference however is that infotaxis can work at the dilute limit, whereas chemotaxis fails [45••].

Qualitatively, it is known that when worms are transferred from plates with food to plates without food the animals perform an intense search of a local area [50] for a limited time. Presumably, animals perceived the food to be located nearby based on their prior experience and previous search outcomes. This 'local search' lasts for approximately 15 min. From a physical perspective, this search problem is similar to the infotaxis navigation considered above. However, an important difference is that in this case there are *no* odorant cues. Therefore, one might expect that all of the animal's behavior must be guided by the dynamics implied by its prior beliefs summarized by $P(r_{current})$. Surprisingly, the infotaxis solution in this context exhibits an abrupt transition between a local search phase and a global search phase (Figure 1a), provided we allow for the possibility that the source is not located within the modeled area (the full extent of the distribution in Figure 1a). The corresponding probability $p(t) = A$ evaluated at time t is updated in a Bayesian manner for the next time step: $p_{t+1}(A|0) = p_t(0|A)p_t(A)/p_t(0)$. In the beginning of the

Figure 1
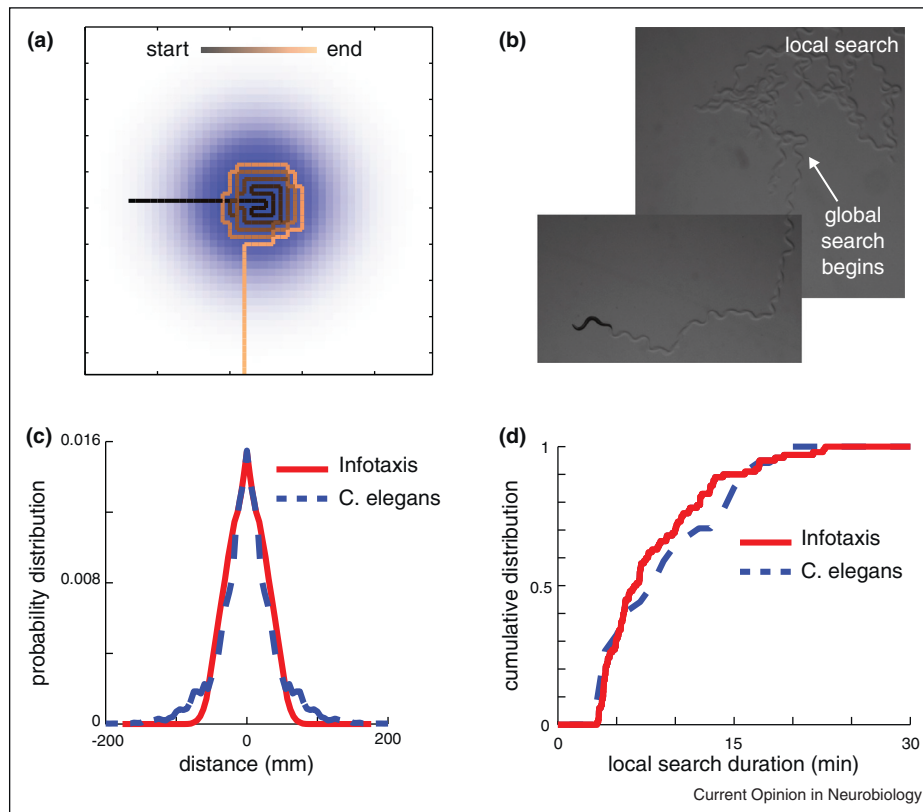
Information maximization explains transition between tasks in the absence of new sensory stimuli. **(a)** Example infotaxis trajectory exhibits a transition from local search to global search. The background color-map shows the initial probability distribution of target locations. Search proceeds following an expectation of odor detection but eventually transitions towards straight motion away from the initial maximum. **(b)** Similar behavior is observed in the case of *C. elegans* once they are removed from a patch of bacteria. The initial trajectory has lots of turns. After approximately 15 min, the animal stops turning and moves away from the local area. **(c)** Information maximization accounts for the spatial distribution of worm positions at the end of the local search. **(d)** Information maximization accounts for the duration of local search in *C. elegans*.

search $p_{t=0}(A) = 1$. The transition from local search to the global search in the model occurs when $p_t(A)$ reaches zero. This transition matches the worm behavior both qualitatively (Figure 1b) and quantitatively in terms of the distribution of worm positions at the end of the local search phase (Figure 1c). Importantly the same set of parameters in the infotaxis model can also account for the duration of the local search (Figure 1d). This match is achieved without further adjustments in the model because the temporal and spatial scales are related by the known speed of worm movements on the plate (0.2 mm/s), which remains unaltered during both the local and global phases of the search [50]. A similar optimal allocation of time on a given task to maximize information was also described by a recent theoretical framework that represented human attention as a decision to interrupt the current task or persevere [51].

The infotaxis model makes a number of important predictions that qualitatively differ from the simpler chemotaxis model of navigation. A chemotaxis model is based on the computation of a gradient. It predicts that the animal's behavioral response will be affected by the magnitude of a drop in odorant concentration. In contrast, the infotaxis model makes predictions based mainly on the relative distribution of food in space rather than on its concentration. If the worms are transferred from plates that have lawns of bacteria (food for *C. elegans*) that have the same size and different concentrations, then the chemotaxis (but not infotaxis) model would predict a stronger initial response from the animal when it is transferred from a more concentrated lawn. Preliminary evidence in these experiments supports the more involved infotaxis model of behavior. A related prediction is that the duration of the search may depend on the size of the animal, because the trajectory should remain the same in dimensionless units where the width of the prior distribution is normalized by the searcher size. Finally, the search duration is dependent on the diffusion properties of the odorant and would suggest that the search duration may depend on the type of bacteria to which the worms were acclimated before removal from food. Verifying these predictions of the infotaxis model would set the precedent that even animals with a relative simple nervous system (302 neurons, in the case of *C. elegans*) perform computations based on a 'mental map' of likely food locations by continuously updating probabilities across the range of spatial locations around them.

Criticisms of information maximization as a behavioral strategy do exist. For example, it has been argued that these strategies do not always guarantee maximum fitness [52]. It is noteworthy that the deviations in optimality when using information maximization have been primarily observed in a dynamic situation where the absolute knowledge of the target position does not guarantee that

this position could be reached before the target would move again. In such cases, perhaps the paradox can be resolved by considering maximizing information not only about the current location of the target but also about its future positions, that is predictive coding [53] and predictive information [4••,54]. Interestingly, this has recently been discussed in the context of neuronal responses. Recent analyses show that synergistic effects across multiple neurons are much stronger when one considers predictive information compared to information about preceding stimuli that caused the neural responses [2••], suggesting that the combinatorial power of neural responses across a population is aimed at maximizing information about future events and not the prior sensory inputs. Furthermore, it is true that reward function that guides the animal's behavior might not correspond directly to information gain, even when integrated to infinite future times. However, to operate as well as possible given the constraints on information acquisition, theoretical arguments indicate that the animal must adjust its internal reward structure to coincide precisely with the predictive information from the past signals about the future trajectories [4••]. This situation was termed as the case 'perfectly adapted environment' in [4••], but perhaps another way of describing this type of adaptation would be to emphasize that it directs adaptation in the reward function of an agent to the statistics of the environment. One would also expect that the two adaptive processes: adaptation of the encoding function to maximize the reward and adaptation of the reward function itself might be achieved over different time scales [4••]. On short time scales, the optimal behavioral strategy is determined by the current reward structure. However, over longer time scales the reward function adapts to coincide with the available predictive information present in stimuli. A recent review summarizes the experimental evidence supporting this hypothesis [55].

Applying the infotactic perspective to adaptive properties in neural systems, it is worth noting that most of the studies of neural adaptation focused on adaptation to time averaged statistical properties of the stimulus such as mean, variance, covariance or perhaps even higher-order moments. Yet sensory neurons also operate in 'dilute' conditions, that is at low signal-to-noise ratios. For example, a recent study shows that the gain of retinal ganglion cells can be affected by detection of a single photon [46•]. It is worth noting that infotaxis is a Bayesian approach based on times of individual (often binary) detection events. Given the success of Bayesian approaches in accounting for neural adaptation based on time averaged stimulus properties, such as mean, variance, and kurtosis [36••], the more nuanced infotaxis approach might provide a new frontier for understanding adaptive functions in the nervous system.

Figure 2
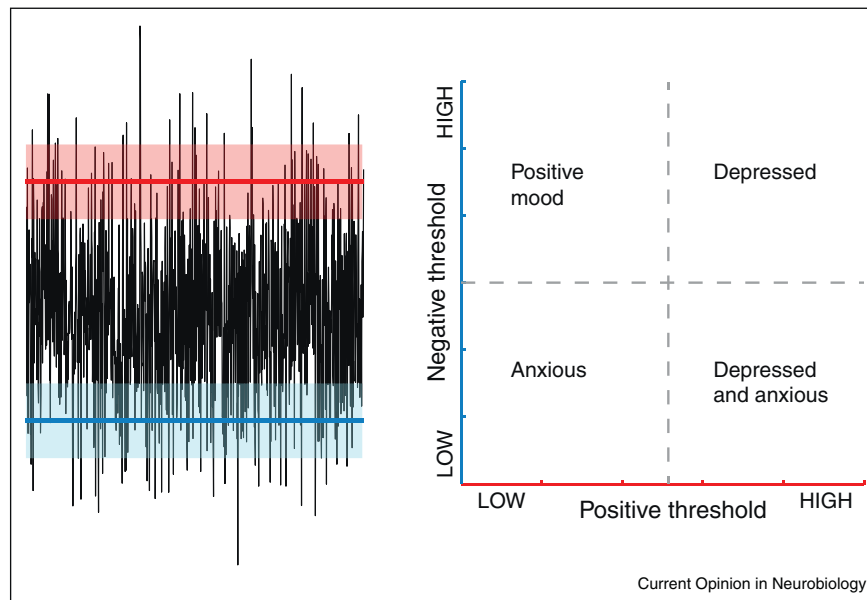
Information theory approach to characterizing mood disorders. **(a)** Separate neural mechanisms for encoding rewards and punishments can be explained using similar information-theoretic arguments that were used in sensory coding to explain the separation into OFF and ON channels. **(b)** Following [56••], variations in positive and negative thresholds correspond to four basic mood states. The setting of these thresholds can be explained by a Bayesian model based on prior outcomes of the animal–environment interactions [56••].

It would be exciting to see if information theory could account for long term adaptive changes in behavior, including mood particularly in the case of humans. There is an emerging view that long lasting mental states, such as depression, anxiety, optimism and pessimism, represent a proper integrative response of an animal to a sequence of events from its prior experiences [56••]. This is an advantageous response from an evolutionary point of view for several reasons. First, many stimuli are ambiguous and lead to reward or punishment with some probability, forcing an animal to make a choice of whether to allocate its effort to pursue a rewarding outcome or to prepare to minimize the consequences of a negative outcome. The threshold for triggering one or the other action cannot be set in a static manner, because the relative costs of false positives and negatives depend on the physical state of the animal and on the state of the environment. Therefore, to make an optimal decision, from a Bayesian point of view [57], the animal has to take into account the outcomes of previous decisions made a recent past, which presumably were characterized by similar states of the animal and its environment. To map this onto mood, a depressed state would then correspond mathematically to having a lower threshold for predicting a negative outcome. It is remarkable that thresholds for betting on positive versus negative outcomes can be changed by similar pharmacological and behavioral modifications in both humans and laboratory rodents [56••]. Furthermore, neural mechanisms underlying emotional states are conserved across a wide variety of organisms, with important homologues between vertebrate and invertebrate species [56••]. In particular, a prominent observation from animal learning studies is that neural mechanisms of reward and punishment are subserved by largely distinct circuits. The analogy to ON-channel and OFF channel in vision might be more than a mere coincidence. Instead, the separate processing of rewards and punishments might reflect the needs to achieve maximally informative coding under metabolic constraints (Figure 2), following the same arguments that were used in vision to explain the existence of separate ON and OFF pathways [58,59•,60–61]. It seems likely that information theory could provide a completely novel and quantitative way of characterizing mood and its disorders.

Unlike in physics, where great successes began with Newton's realization that the physical laws are the same whether on this planet or in space, in molecular biology, which bloomed with the discovery of a universal genetic code, we do not yet have a unifying framework to work with at the systems and behavioral levels. Information theoretic ideas have been successful at explaining properties of the nervous system from individual neurons to populations of neurons, the building blocks of behavior [2••]. Although the information maximization framework has been tested so far on only in a handful of different types of behavior, it seems well suited for providing such a

unifying framework for understanding different facets of behavior, from adaptation in sensory systems to the adaptation of reward circuits on longer time scales so that they can better guide learning.

Acknowledgments

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. Hawking S: *Fear the Aliens*. Discovery Channel; 2010.

2. Bialek W: *Biophysics Searching for Principles*. 2013.
An encompassing review of current ideas based on information theory and the function of living systems, from protein folding to neural transmission to infotaxis models.

3. Polani D: **Information: currency of life?** *HFSP J* 2009, **3**:307-316.

4. Tishby N, Polani D (Eds): *Information Theory of Decisions and Actions*. New York: Springer; 2011.
This paper describes how one can compute the cumulated information processing cost expected for future trajectories that start from a given state. The corresponding quantity termed 'information-to-go' obeys a self-consistent Bellman-type equation. Finally, the authors describe the case of 'perfectly adapted environments' where reward function equals predictive information between past and future stimuli.

5. Bray D: **Protein molecules as computational elements in living cells**. *Nature* 1995, **376**:307-312.

6. Cardoso MM, Sirotin YB, Lima B, Glushenkova E, Das A: **The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components**. *Nat Neurosci* 2012, **15**:1298-1306.
A thought provocative study demonstrating that a substantial part of the signals measuring changes in the blood oxygenation level are anticipatory and task-related, as opposed to causally elicited by prior sensory stimuli.

7. Powell WB: *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley; 2011.

8. Wark B, Lundstrom BN, Fairhall A: **Sensory adaptation**. *Curr Opin Neurobiol* 2007, **17**:423-429.

9. Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME: **Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex**. *PLoS Biol* 2007, **5**:e19.

10. Liu YS, Stevens CF, Sharpee TO: **Predictable irregularities in retinal receptive fields**. *Proc Natl Acad Sci U S A* 2009, **106**:16499-16504.

11. Fitzgerald JD, Sharpee TO: **Maximally informative pairwise interactions in networks**. *Phys Rev E Stat Nonlin Soft Matter Phys* 2009, **80**:031914.

12. Tkacik G, Prentice JS, Balasubramanian V, Schneidman E: **Optimal population coding by noisy spiking neurons**. *Proc Natl Acad Sci U S A* 2010, **107**:14419-14424.

13. Brenner N, Bialek W, de Ruyter van Steveninck R: **Adaptive rescaling maximizes information transmission**. *Neuron* 2000, **26**:695-702.

14. Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR: **Efficiency and ambiguity in an adaptive neural code**. *Nature* 2001, **412**:787-792.

15. Rieke F, Rudd ME: **The challenges natural images pose for visual adaptation**. *Neuron* 2009, **64**:605-616.

16. Yaron A, Hershenhoren I, Nelken I: **Sensitivity to complex statistical regularities in rat auditory cortex**. *Neuron* 2012, **76**:603-615.

17. Kompaniez E, Sawides L, Marcos S, Webster MA: **Adaptation to interocular differences in blur**. *J Vis* 2013, **13**:19.

18. Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD: **Adaptive filtering enhances information transmission in visual cortex**. *Nature* 2006, **439**:936-942.

19. Webster MA, MacLeod DI: **Visual adaptation and face perception**. *Philos Trans R Soc Lond B Biol Sci* 2011, **366**:1702-1725.

20. Keefe BD, Dzhelyova M, Perrett DI, Barraclough NE: **Adaptation improves face trustworthiness discrimination**. *Front Psychol* 2013, **4**:358.

21. Sharpee TO, Bialek W: **Neural decision boundaries for maximal information transmission**. *PLoS ONE* 2007.

22. Horwitz GD, Hass CA: **Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing**. *Nat Neurosci* 2012, **15**:913-919.
This study demonstrates that differences between decision boundaries of simple and complex cells extend to the color domain, with simple cells characterized by extended contours and complex cells characterized by closed decision boundary contours.

23. Rust NC, Schwartz O, Movshon JA, Simoncelli EP: **Spatiotemporal elements of macaque v1 receptive fields**. *Neuron* 2005, **46**:945-956.

24. Shamir M, Sompolinsky H: **Nonlinear population codes**. *Neural Comput* 2004, **16**:1105-1136.

25. Sompolinsky H, Yoon H, Kang K, Shamir M: **Population coding in neuronal systems with correlated noise**. *Phys Rev E Stat Nonlin Soft Matter Phys* 2001, **64** 04 0519.

26. Averbeck BB, Latham PE, Pouget A: **Neural correlations, population coding and computation**. *Nat Rev Neurosci* 2006, **7**:358-366.

27. Gutnisky DA, Dragoi V: **Adaptive coding of visual information in neural populations**. *Nature* 2008, **452**:220-224.

28. Wang Y, Iliescu BF, Ma J, Josic K, Dragoi V: **Adaptive changes in neuronal synchronization in macaque V4**. *J Neurosci* 2011, **31**:13204-13213.

29. Jeanne JM, Sharpee TO, Gentner TQ: **Associative learning enhances population coding by inverting interneuronal correlation patterns**. *Neuron* 2013, **78**:352-363.

30. Rothschild G, Cohen L, Mizrahi A, Nelken I: **Elevated correlations in neuronal ensembles of mouse auditory cortex following parturition**. *J Neurosci* 2013, **33**:12851-12861.

31. Mortimer D, Feldner J, Vaughan T, Vetter I, Pujic Z, Rosoff WJ, Burrage K, Dayan P, Richards LJ, Goodhill GJ: **Bayesian model predicts the response of axons to molecular gradients**. *Proc Natl Acad Sci U S A* 2009, **106**:10296-10301.

32. Mortimer D, Dayan P, Burrage K, Goodhill GJ: **Bayes-optimal chemotaxis**. *Neural Comput* 2011, **23**:336-373.

33. Soo FS, Schwartz GW, Sadeghi K, Berry MJ 2nd: **Fine spatial information represented in a population of retinal ganglion cells**. *J Neurosci* 2011, **31**:2145-2155.

34. Doi E, Gauthier JL, Field GD, Shlens J, Sher A, Greschner M, Machado TA, Jepson LH, Mathieson K, Gunning DE et al.: **Efficient coding of spatial information in the primate retina**. *J Neurosci* 2012, **32**:16256-16264.

35. Gauthier JL, Field GD, Sher A, Greschner M, Shlens J, Litke AM, Chichilnisky EJ: **Receptive fields in primate retina are coordinated to sample visual space more uniformly**. *PLoS Biol* 2009, **7**:e0631000.

36. Wark B, Fairhall A, Rieke F: **Timescales of inference in visual**
•• **adaptation**. *Neuron* 2009, **61**:750-761.
This study combines experiment and theory to show that adaptation time scales are optimal in the sense that they track accumulation of statistical evidence towards a change in parameters.

37. Bonin V, Mante V, Carandini M: **The statistical computation underlying contrast gain control**. *J Neurosci* 2006, **26**:6346-6353.

38. Mease RA, Famulare M, Gjorgjieva J, Moody WJ, Fairhall AL: **Emergence of adaptive computation by single neurons in the developing cortex**. *J Neurosci* 2013, **33**:12154-12170.

39. Hong S, Lundstrom BN, Fairhall AL: **Intrinsic gain modulation and adaptive neural coding**. *PLoS Comput Biol* 2008, **4**:e1191000.

40. Borst A, Flanagin VL, Sompolinsky H: **Adaptation without parameter change: dynamic gain control in motion detection**. *Proc Natl Acad Sci U S A* 2005, **102**:6172-6176.

41. Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA: **A**
• **functional and perceptual signature of the second visual area in primates**. *Nat Neurosci* 2013, **16**:974-981.
This article describes a set of results showing differences in neural responses in the primary and secondary visual areas to stimuli with and without pairwise correlations between the output Gabor-like filters across different positions, orientations, and scale.

42. McDermott JH, Schemitsch M, Simoncelli EP: **Summary**
• **statistics in auditory perception**. *Nat Neurosci* 2013, **16**:493-498.
Using sound textures the authors show that auditory perception is based on time-averaged statistics in within (approximately) spectrotemporal domain. This conclusion is based on the observation that discrimination between textures with different pairwise statistics improves with time, whereas discrimination between textures with the same statistics deteriorates with time. This is presumably due to different examples of textures with the same pairwise statistics need to be discriminated based on higher order statistics that are not actually used in discrimination.

43. Victor JD, Thengone DJ, Conte MM: **Perception of second- and**
• **third-order orientation signals and their interactions**. *J Vis* 2013, **13**:21.
An interesting account of how second-order and higher order statistics of visual patterns affect visual perception.

44. Butler CG: **Insect pheromones**. *Biol Rev* 1967, **42**:42-87.

45. Vergassola M, Villermaux E, Shraiman BI: **'Infotaxis' as a strategy**
•• **for searching without gradients**. *Nature* 2007, **445**:406-409.
This paper shows that one can successfully find targets by following the gradient of information rather than the concentration gradient. Close to the odorant source, the two search strategies coincide. However, far from the source where odorant detection events are rare and concentration gradients are unreliable, the information gradient can successfully guide the searcher towards its goal.

46. Schwartz GW, Rieke F: **Controlling gain one photon at a time**.
• *Elife* 2013, **2**:e70046.
This paper demonstrates how the gain of retinal ganglion cells can be sensitive to individual photo absorption events.

47. Najemnik J, Geisler WS: **Optimal eye movement strategies in**
•• **visual search**. *Nature* 2005, **434**:387-391.
This paper describes that human eye movements approximate performance based on an optimal Bayesian model where the distribution of target location is updated following each fixation. This approach can be mapped onto the infotaxis model and predicts many interesting properties of human eye movements, including elimination saccades and center-of-gravity saccades.

48. Rao RP, Zelinsky GJ, Hayhoe MM, Ballard DH: **Eye movements in iconic visual search**. *Vision Res* 2002, **42**:1447-1463.

49. Lewis RL, Shvartsman M, Singh S: **The adaptive nature of eye movements in linguistic tasks: how payoff and architecture shape speed-accuracy trade-offs**. *Top Cogn Sci* 2013, **5**:581-610.

50. Chalasani SH, Chronis N, Tsunozaki M, Gray JM, Ramot D, Goodman MB, Bargmann CI: **Dissecting a circuit for olfactory behaviour in** *Caenorhabditis elegans*. *Nature* 2007, **450**:63-70.

51. Ballard DH, Kit D, Rothkopf CA, Sullivan B: **A hierarchical modular architecture for embodied cognition**. *Multisens Res* 2013, **26**:177-204.

52. Agarwala EK, Chiel HJ, Thomas PJ: **Pursuit of food versus pursuit of information in a Markovian perception-action loop model of foraging**. *J Theor Biol* 2012, **304**:235-272.

53. Rao RP, Ballard DH: **Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects**. *Nat Neurosci* 1999, **2**:79-87.

54. Bialek W, Nemenman I, Tishby N: **Predictability, complexity, and learning**. *Neural Comput* 2001, **13**:2409-2463.

55. Gottlieb J, Oudeyer PY, Lopes M, Baranes A: **Information-seeking, curiosity, and attention: computational and neural mechanisms**. *Trends Cogn Sci* 2013, **17**:585-593.

56. Nettle D, Bateson M: **The evolutionary origins of mood and its**
•• **disorders**. *Curr Biol* 2012, **22**:R712-R721.
This review explains how depression and anxiety have positive survival value based on signal detection theory. The conclusions may have deep implications for societal policies aimed at reducing and mitigating the effects of mood disorders.

57. Jaynes ET: *Probability Theory: The Logic of Science*. Cambridge, England: Cambridge University Press; 2003, .

58. Balasubramanian V, Sterling P: **Receptive fields and functional architecture in the retina**. *J Physiol* 2009, **587**:2753-2767.

59. Ratliff CP, Borghuis BG, Kao YH, Sterling P, Balasubramanian V:
• **Retina is structured to process an excess of darkness in natural scenes**. *Proc Natl Acad Sci U S A* 2010, **107**:17368-17373.
This paper presents an intriguing observation that there are more negative than positive contrasts in natural scenes. This makes is possible to quantitatively explain the observed differences in the density of OFF and ON mosaics of retinal ganglion cells.

60. Kastner D, Baccus S, Sharpee TO: **Optimal placement of dynamic range by coordinated populations of retinal ganglion cells**. *Cosyne Abstracts*; *Salt Lake City, USA: 2012*.

61. Gjorgjieva J, Sompolinsky H, Meister M: **Parallel pathways for information processing in the retina: the ON and OFF dichotomy**. *Cosyne Abstracts*; *Salt Lake City, USA: 2012*.