allowed investigators to link specific behaviors to identified neurons in a causal fashion[10–13]. Botta *et al.*[1] take this many steps further and now show a receptor subunit modifying the activity of a specific subset of neurons that encodes specific information that promotes anxiogenic behavior.

Like most ground-breaking discoveries, these observations also raise intriguing questions. For example, what is the mechanism through which this system translates the pairing between CS and unconditioned stimulus into a decrease in the tonic current? Is it a result of specific patterns of activity in afferent inhibitory neurons? Is there a change in glial uptake of GABA that causes a desensitization of extrasynaptic $GABA_A$ receptors? Is a specific neuromodulator involved? Along these lines, a recent report has shown that glucocorticoids are necessary for the decrease in I_{tonic} in amygdala neurons following chronic stress[14]. In addition, the opposing changes in I_{tonic} described in PKCδ$^-$ neurons are intriguing. Previous work from this group and other has shown that PKCδ$^+$ neurons in the lateral region of the central nucleus of the amygdala (CEA) also receive local inhibitory inputs from PKCδ$^-$ neurons[2,3], so the concerted effects of these reciprocal changes on the output of this microcircuit remain unresolved.

Finally, others have reported that these same neurons, when optogenetically activated, inhibit feeding and are anxiolytic, not anxiogenic[15]. How does one reconcile these opposing observations? One possibility is that PKCδ$^+$ neurons do not comprise a homogeneous population with respect to output targets. Another is that the mode of photoactivation is important. Botta *et al.*[1] used a constant pulse of blue light for 30 s, whereas light was delivered at 5 Hz in ref. 15. The latter could induce synchronization of a population of PKCδ$^+$ neurons, but, according to Botta *et al.*[1], pulsed light stimulation is less effective at increasing firing rate. The reasons for this are unclear and suggest that there are many more issues to resolve.

The implications of understanding the mechanistic underpinnings of fear generalization are obvious for conditions such as post-traumatic stress disorder, but these findings highlight a distinct role for tonic GABA currents and may also provide important clues about how differences in receptor subunits may predispose some individuals to be more or less sensitive to becoming anxious. In addition, the approach used here provides a template for other investigators to causally link specific receptors and proteins in defined cell populations in the brain to defined behaviors.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Botta, P. *et al. Nat. Neurosci.* **18**, 1493–1500 (2015).
2. Ciocchi, S. *et al. Nature* **468**, 277–282 (2010).
3. Haubensak, W. *et al. Nature* **468**, 270–276 (2010).
4. De Oca, B.M., DeCola, J.P., Maren, S. & Fanselow, M.S. *J. Neurosci.* **18**, 3426–3432 (1998).
5. Farrant, M. & Nusser, Z. *Nat. Rev. Neurosci.* **6**, 215–229 (2005).
6. Anstee, Q.M. *et al. Nat. Commun.* **4**, 2816 (2013).
7. Zurek, A.A. *et al. J. Clin. Invest.* **124**, 5437–5441 (2014).
8. Maguire, J. & Mody, I. *Neuron* **59**, 207–213 (2008).
9. Janak, P.H. & Tye, K.M. *Nature* **517**, 284–292 (2015).
10. Yiu, A.P. *et al. Neuron* **83**, 722–735 (2014).
11. Tye, K.M. *et al. Nature* **471**, 358–362 (2011).
12. Namburi, P. *et al. Nature* **520**, 675–678 (2015).
13. Lammel, S. *et al. Nature* **491**, 212–217 (2012).
14. Liu, Z.-P. *et al. Mol. Brain* **7**, 32 (2014).
15. Cai, H., Haubensak, W., Anthony, T.E. & Anderson, D.J. *Nat. Neurosci.* **17**, 1240–1248 (2014).

Explaining the especially pink elephant

Jonathan W Pillow

A new study shows that an efficient allocation of sensory resources can lead to Bayesian estimates that are biased away from the prior, accounting for effects such as the bias toward oblique angles in orientation perception.

Two hallmark ideas of theoretical neuroscience are efficient coding and Bayesian perception. The first of these ideas says roughly that sensory systems should allocate their resources to maximize information about the environment: do not waste space building detectors for pink elephants, blue mice or other fanciful beings that are unlikely to turn up very often in real life. The second idea says (also very sensibly) that when asked to make perceptual judgments we should combine information from our senses with prior beliefs about the world: in conditions of good visibility we should trust our eyes, but when visibility is poor and our eyes report something unexpected—for example, a pink elephant in the room—we should rely more heavily on top-down, 'prior' information. Although these ideas have had major roles in explaining properties of sensory neural responses and perceptual behavior, they have not previously been considered in a single framework. A study by Wei and Stocker[1] in this issue of *Nature Neuroscience* bridges this gap. The authors show that, remarkably, an observer governed by both principles may report seeing an especially pink elephant when shown only a slightly rosy one.

The study refers to such percepts as 'anti-Bayesian' because they involve the percept of something that is less probable under the prior (which holds that pink elephants are unlikely) than the thing actually presented. This result is surprising and seems highly counterintuitive. The authors show not only that it holds mathematically, but that it can account for a variety of published effects involving the perception of orientation and spatial frequency[1].

The starting point for this work is Attneave's and Barlow's 'efficient coding' or 'redundancy reduction' hypothesis[2,3]. This idea, which has guided neuroscience over the past five decades, seeks to understand the design principles governing sensory neurons using information theory. The basic theory states that neural responses should convey maximal information about the environment. Formally, the neural encoding distribution $p(r \mid \theta)$, which describes how a stimulus θ is transformed into a noisy neural response r, should be set up to maximize the mutual information between θ and r. Of course, this depends on $p(\theta)$, the prior distribution over stimuli in the natural environment.

Recent work has sought to attack this problem using Fisher information, a quantity that can be used to approximate mutual information and can be calculated from neural tuning curves. A powerful recent result is that an efficient code (that is, one conveying maximal information) can be obtained by setting Fisher information $J(\theta)$ proportional to $p^2(\theta)$, the square of the prior distribution[4,5]. In essence, the allocation of neural resources (as defined by Fisher information) should be even more concentrated than the prior. To illustrate the idea, a bell-shaped prior distribution (**Fig. 1a**) defines an optimal Fisher information curve for a neural population, which can be achieved

Jonathan W. Pillow is at the Princeton Neuroscience Institute, Department of Psychology and Center for Statistics and Machine Learning at Princeton University, Princeton, New Jersey, USA.
e-mail: pillow@princeton.edu

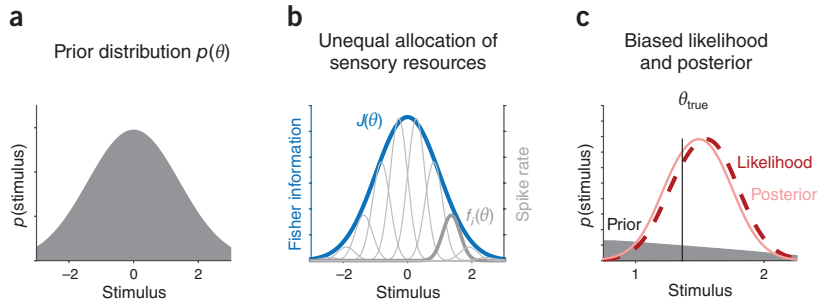

Figure 1 Efficient neural coding and Bayesian inference. (**a**) Example stimulus prior distribution $p(\theta)$. (**b**) The optimal encoding should have Fisher information $J(\theta)$ proportional to the prior squared (blue trace), which can be achieved with a population of Poisson neurons with Gaussian-shaped tuning curves (gray traces). (**c**) A stimulus $\theta_{true} = 1.4$ elicits a single spike from the neuron with tuning curve centered at $\theta_i = 1.4$ (thick gray trace in **b**). The resulting likelihood (dashed line) is shifted rightwards from θ_{true} as a result of a term involving the negative sum of the tuning curves. The posterior distribution (pink) over θ given the response, given by the normalized product of prior and likelihood, is slightly left of the likelihood but still biased rightwards relative to θ_{true}. Wei and Stocker[1] show that this bias is a generic feature of codes with Fisher information concentrated as in **b**. Vertical scales in **a** and **c** are different; tickmarks reflect the same intervals.

by Poisson neurons with appropriately scaled tuning curves (**Fig. 1b**). For this population, the most probable stimuli (those near $\theta = 0$) will elicit many spikes, whereas those in the tails will elicit few spikes, in accordance with our intuition that an efficient code should allocate few sensory resources to pink elephants.

The second fundamental component of the work by Wei and Stocker[1] is the theory of perception as Bayesian inference. The core idea that perceptual inference depends on a combination of 'bottom-up' sensory information and 'top-down' prior information dates back at least to Helmholtz, and Bayesian theories have a rich and successful recent history in perceptual psychology and neuroscience[6–10]. Formally, the theory asserts that a percept depends entirely on a posterior distribution $p(\theta \mid r)$, which (according to the beloved formula known as Bayes' rule) is proportional to $p(r \mid \theta) \times p(\theta)$, the likelihood times the prior. The likelihood $p(r \mid \theta)$ captures bottom-up sensory information that an observed neural response r carries about the (unknown) stimulus θ, whereas the prior $p(\theta)$ represents top-down beliefs about θ. The posterior summarizes the observer's final state of knowledge after combining the sensory measurements r with information from the prior. In Bayesian analyses, the posterior is always biased toward the prior relative to the likelihood (**Fig. 1c**).

Conceptually, the theories of efficient coding and Bayesian perception fit together naturally. They make complementary demands on a neural population: efficient coding wants the encoding function $p(r \mid \theta)$ to maximize information about stimuli drawn from $p(\theta)$, whereas Bayesian perception wants decoding to rely on the posterior, formed by the product of $p(r \mid \theta)$ and $p(\theta)$. How then can an efficient Bayesian code yield percepts that are systematically biased away from the prior? In other words, why does the natural synthesis have such bizarre consequences?

An intuitive explanation for the existence of anti-Bayesian percepts is that—in an efficient code—sensory responses provide relatively little information about the low-prior-probability regions of stimulus space. In essence, the likelihood $p(r \mid \theta)$ typically cannot rule out values of θ for which $p(\theta)$ is low as easily as it can rule out values of θ for which $p(\theta)$ is high, simply because there are not many neural resources devoted to regions where $p(\theta)$ is low. This skew in the likelihood (it falls off more steeply on one side than the other) is inherited by the posterior, and the mean of the posterior, which Wei and Stocker[1] propose as the basis for observers' perceptual judgments, is therefore biased away from high-prior-probability regions of stimulus space (**Fig. 1c**).

Wei and Stocker[1] formalize this intuition and show that the phenomenon holds (under certain regularity conditions) for other neural populations with the requisite allocation of Fisher information, regardless of what tuning curve shapes are used to achieve it. More importantly, they show that the phenomenon arises in human observers' judgments of orientation and spatial frequency. Observers presented with an orientated stimulus offset from a cardinal orientation (0° or 90°) perceive it as closer to the oblique orientation (45° or 135°) (ref. 11)—that is, away from the natural prior distribution over orientations. Similarly, observers systematically misperceive a Gabor grating patch as biased toward higher spatial frequencies, away from the $1/f$ distribution of natural images that favors low frequencies[12]. These phenomena, which seem to contradict the fundamental principle that Bayesian estimates are biased toward the prior, are perfectly consistent with the Bayesian model put forth by Wei and Stocker[1]—a remarkable finding.

Of course, the old rules and intuitions around Bayes have not all been suspended. In the proposed framework, the posterior still lives in between prior and likelihood, and the Bayesian estimate is still more biased toward the prior than the mean of the (normalized) likelihood. The key culprit is the likelihood, which can behave in unexpected ways when uncertainty is distributed unequally[13]. Here efficient population codes have the surprising property that the likelihood tends to skew away from the prior, a property that the posterior simply inherits.

Wei and Stocker's[1] work represents the most promising attempt to date to reconcile Bayesian perceptual inference with the existence of anti-Bayesian psychophysical biases, but it nevertheless has several possible limitations. For one, it relies on loss functions for encoding and decoding that are slightly mismatched. The assumed encoder maximizes mutual information, whereas the assumed decoder minimizes mean-squared error; if these choices are made consistent, anti-Bayesian effects may not arise. For another, Fisher information provides only an asymptotic approximation to mutual information, and the optimal code for high noise levels may differ[14,15]. Nevertheless, Wei and Stocker's study[1] represents a new and important conceptual advance that illuminates the unexpected power of Bayesian models to explain seemingly non-Bayesian perceptual phenomena. It provides an elegant unification of efficient coding and Bayesian inference, and offers a new explanation for the tradeoffs between attraction and repulsion from the prior that will stimulate new directions of experimental and theoretical research.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Wei, X. & Stocker, A.A. *Nat. Neurosci.* **18**, 1509–1517 (2015).
2. Attneave, F. *Psychol. Rev.* **61**, 183–193 (1954).
3. Barlow, H.B. in *Sensory Communication* (ed. Rosenblith, W.A.) 217–234 (MIT Press, 1961).
4. Brunel, N. & Nadal, J.P. *Neural Comput.* **10**, 1731–1757 (1998).
5. Ganguli, D. & Simoncelli, E.P. *Neural Comput.* **26**, 2103–2134 (2014).
6. Knill, D. & Richards, W. *Perception as Bayesian Inference* (Cambridge University Press, 1996).
7. Geisler, W.S., Perry, J., Super, B. & Gallogly, D. *Vision Res.* **41**, 711–724 (2001).
8. Ernst, M.O. & Banks, M.S. *Nature* **415**, 429–433 (2002).
9. Weiss, Y., Simoncelli, E.P. & Adelson, E. *Nat. Neurosci.* **5**, 598–604 (2002).
10. Körding, K.P. & Wolpert, D. *Nature* **427**, 244–247 (2004).
11. de Gardelle, V., Kouider, S. & Sackur, J. *J. Vis.* **10**, (2010).
12. Georgeson, M.A. & Ruddock, K.H. *Phil. Trans. R. Soc. Lond. B* [and discussion] **290**, 11–22 (1980).
13. Stocker, A.A. & Simoncelli, E.P. *Adv. Neural Inf. Process. Syst.* **18**, 1289 (2006).
14. Bethge, M., Rotermund, D. & Pawelzik, K. *Neural Comput.* **14**, 2317–2351 (2002).
15. Berens, P., Ecker, A., Gerwinn, S., Tolias, A. & Bethge, M. *Proc. Natl. Acad. Sci. USA* **108**, 4423 (2001).