Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).

For many decades, neuroscientists under- stood the brain as a 'stimulus-response' organ, consisting of individual neurons that lie dormant until stimulated. In this traditional model, learning and experience merely modulate neural activity that is driven by sensory events in the world. In recent years, scientists have come to realize that the brain probably does not work this way. Instead, research and theory are con- verging on the idea of the brain as an active inference generator that functions according to a **Bayesian Approach to Probability** :sensory inputs constrain estimates of prior probability (from past experience) to create the posterior probabilities that serve as beliefs about the causes of such inputs in the present.

Brains, it has recently been argued, are essentially **prediction machines**. They are bundles of cells that support perception and action by *constantly attempting to match incoming sensory inputs with top-down expectations or predictions*. This is achieved using a hierarchical *generative model that aims to minimize prediction error* within a bidirectional cascade of cortical processing. Such accounts offer a unifying model of perception and action, illuminate the functional role of attention, and may neatly capture the special contribution of cortical processing to adaptive success. This target article critically examines this "hierarchical prediction machine" approach, concluding that it offers the best clue yet to the shape of a unified science of mind and action..

The task of the brain, when viewed from a certain distance, can seem impossible: it must discover information about the likely causes of impinging signals without any form of direct access to their source. Thus, consider a black box taking inputs from a complex external world. The box has input and output channels along which signals flow. But all that it "knows", in any direct sense, are the ways its own states (e.g., spike trains) flow and alter. In that (restricted) sense, all the system has direct access to is its own states. The world itself is thus off-limits (though the box can, importantly, issue motor commands and await developments). **The brain is one such black box.**

Perception thus involves "explaining away" the driving (incoming) sensory signal by matching it with a cascade of predictions pitched at a variety of spatial and temporal scales. These predictions reflect what the system already knows about the world (including the body) and the uncertainties associated with its own processing. Perception here becomes "theory-laden" in at least one (rather specific) sense: What we perceive depends heavily upon the set of priors (including any relevant hyper-priors) that the brain brings to bear in its best attempt to predict the current sensory signal.

Neuropsychopharmacology REVIEWS (2011) 36, 294–315

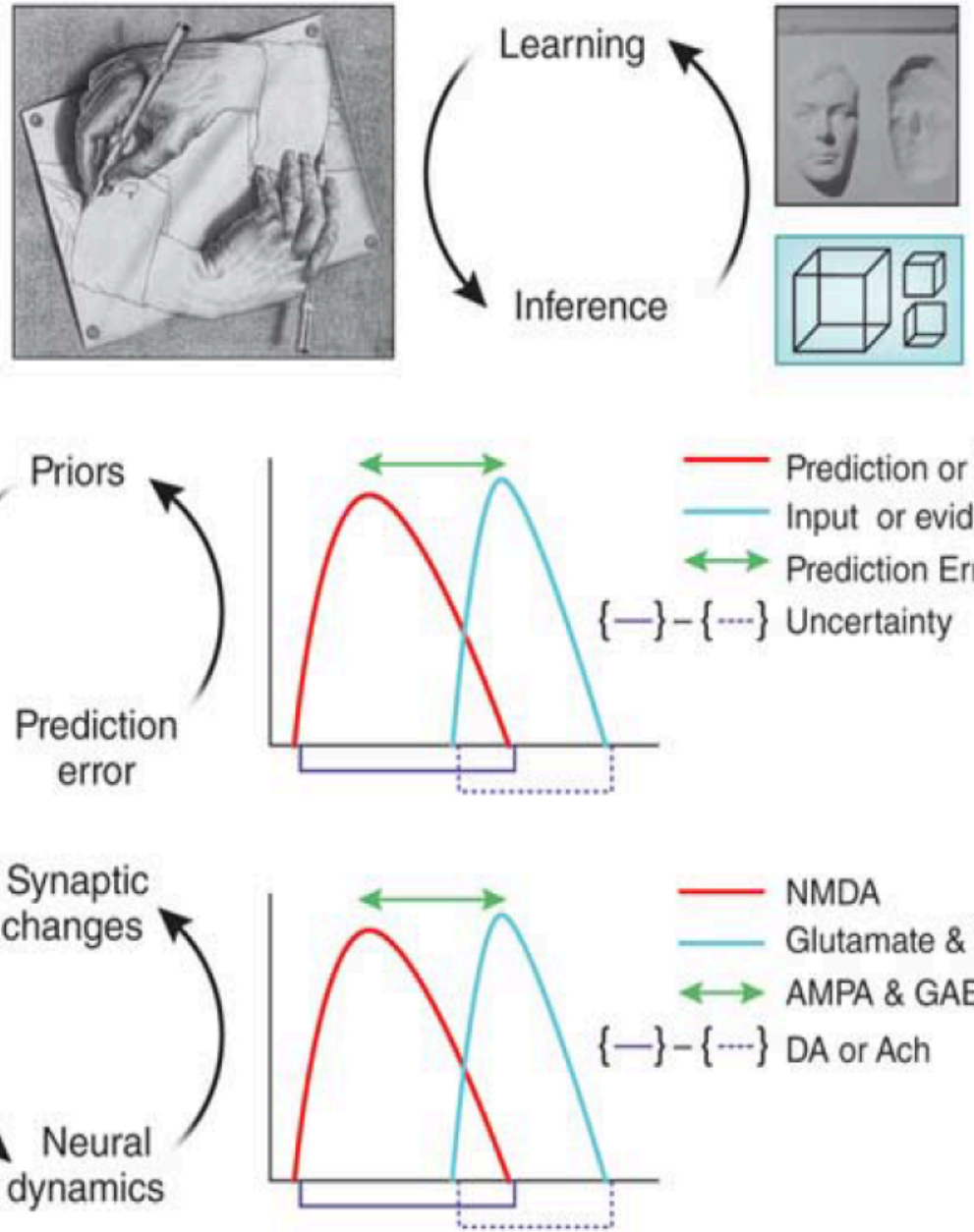
Glutamatergic Model Psychoses: Prediction Error, Learning, and Inference

Why should a neuropsychopharmacologist care about such an arcane mathematical theory? Bayes theorem might help us, as scientists, to reason about new data and evaluate our *a priori* hypotheses: according to Bayes rule, the probability that a hypothesis is true given the observed data (the posterior probability) is proportional to the likelihood (the probability of those data given that the hypothesis is true) multiplied by the prior probability (the probability that the hypothesis was correct before the data were seen). For simplicity, the 'prior probability' is often abbreviated as the '*prior*'.

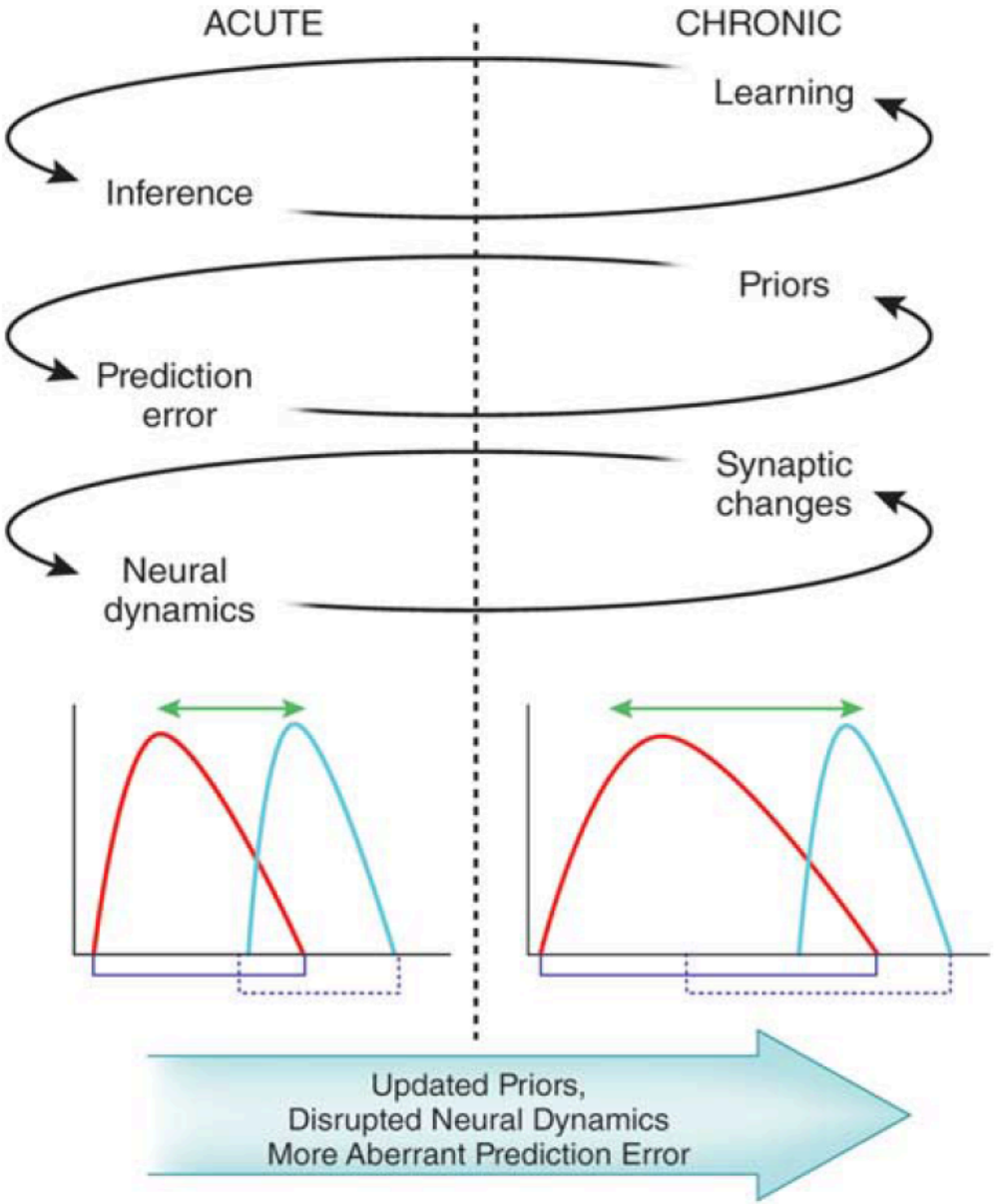
More formally, we can express the relationship between a hypothesis and some new data as:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

Where H is the hypothesis and D is the data. P(H) is the prior probability of H: the probability that H is correct before the data D were observed. P(D|H) is the probability off seeing the data D given that the hypothesis (H) is true. P(D|H) is the likelihood. P(D) is the probability of observing the data D and P(H|D) is the posterior probability.

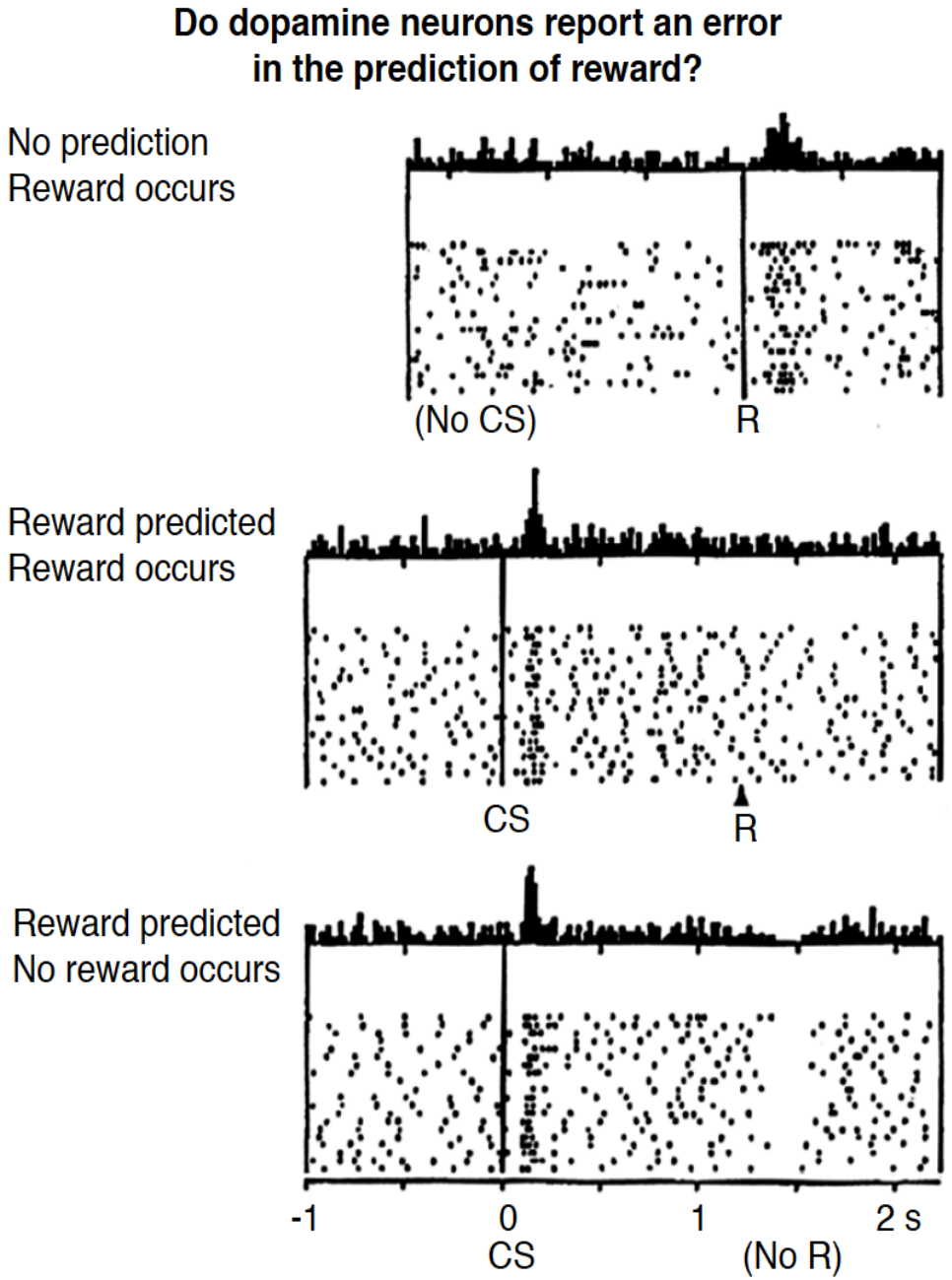


A model of the reciprocal relationships between inference and learning, priors and prediction error, synaptic plasticity and neural dynamics. Inference is encapsulated in the bistable percepts of the Necker Cube, that is, when faced with ambiguous inputs, the brain entertains multiple hypotheses and makes an inference as to the best candidate. The powerful effect of learning on perception is captured by the **hollow mask illusion**, wherein, as a result of our overwhelming experience with faces as convex, we perceive a hollow, concave, inverted mask as convex. *All predictions, or hypotheses that we entertain, have a likelihood distribution, which we compare with the inputs, computing: a prediction error; a degree of uncertainty associated with that prediction error.* We speculate that fast neurotransmitters (GABA and glutamate) may code the prediction error and slower neuromodulators (eg, dopamine and acetylcholine, depending on the task and underlying circuitry) may compute the uncertainty.



The putative effects of acute and chronic ketamine treatment within the Bayesian model. We predict that, with repeated ketamine exposure, aberrant learning (due to deranged synaptic plasticity) and subsequent inappropriate inferences (based on perturbed neural dynamics) lead to maladaptive and inaccurate representations of the world; delusional beliefs.

In recent work, effects of the neurotransmitter dopamine are presented as one possible neural mechanism for encoding precision (see Fletcher & Frith [2009, pp. 53–54] who refer the reader to work on prediction error and the mesolimbic dopaminergic system such as Holleman & Schultz 1998; Waelti et al. 2001). Greater precision (however encoded) means less uncertainty, and is reflected in a higher gain on the relevant error units (see Friston 2005; 2010; Friston et al. 2009). Attention, if this is correct, is simply one means by which certain error-unit responses are given increased weight, hence becoming more apt to drive learning and plasticity, and to engage compensatory action.



Changes in dopamine neurons' output code for an error in the prediction of appetitive events. (**Top**) Before learning, a drop of appetitive fruit juice occurs in the absence of prediction—hence a positive error in the prediction of reward. The dopamine neuron is activated by this unpredicted occurrence of juice. **(Middle**) After learning, the conditioned stimulus predicts reward, and the reward occurs according to the prediction—hence no error in the prediction of reward. The dopamine neuron is activated by the reward-predicting stimulus but fails to be activated by the predicted reward (right). **(Bottom**) After learning, the conditioned stimulus predicts a reward, but the reward fails to occur because of a mistake in the behavioral response of the monkey. The activity of the dopamine neuron is depressed exactly at the time when the reward would have occurred. The depression occurs more than 1 s after the conditioned stimulus without any intervening stimuli, revealing an internal representation of the time of the predicted reward. Neuronal activity is aligned on the electronic pulse that drives the solenoid valve delivering the reward liquid (top) or the onset of the conditioned visual stimulus (middle and bottom). Each panel shows the perievent time histogram and raster of impulses from the same neuron. Horizontal distances of dots correspond to real-time intervals. Each line of dots shows one trial. Original sequence of trials is plotted from top to bottom. CS, conditioned, reward-predicting stimulus; R, primary reward. **Science 275:1593-1599 (1997**).

Andy Clark
School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, EH8 9AD Scotland, United Kingdom
andy.clark@ed.ac.uk
http://www.philosophy.ed.ac.uk/people/full-academic/andy-clark.html

Whatever next? Predictive brains, situated agents, and the future of cognitive science

BEHAVIORAL AND BRAIN SCIENCES (2013), Page 1 of 73doi:10.1017/S0140525X12000477

Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons

Christopher D. Fiorillo,* Philippe N. Tobler, Wolfram Schultz

Uncertainty is critical in the measure of information and in assessing the accuracy of predictions. It is determined by probability P, being maximal at $P = 0.5$ and decreasing at higher and lower probabilities. Using distinct stimuli to indicate the probability of reward, we found that the phasic activation of dopamine neurons varied monotonically across the full range of probabilities, supporting past claims that this response codes the discrepancy between predicted and actual reward. In contrast, a previously unobserved response covaried with uncertainty and consisted of a gradual increase in activity until the potential time of reward. The coding of uncertainty suggests a possible role for dopamine signals in attention-based learning and risk-taking behavior.

Science 299, 1898 (2003);
Christopher D. Fiorillo et al.
Discrete Coding of Reward Probability and Uncertainty by Dopamine

NATURE | VOL 427 :244-247| (2004)

Bayesian integration in sensorimotor learning

Konrad P. Körding & Daniel M. Wolpert

Sobell Department of Motor Neuroscience, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK

When we learn a new motor skill, such as playing an approaching tennis ball, both our sensors and the task possess variability. Our sensors provide imperfect information about the ball's velocity, so we can only estimate it. Combining information from multiple modalities can reduce the error in this estimate[1–4]. On a longer time scale, not all velocities are a priori equally probable, and over the course of a match there will be a probability distribution of velocities. According to bayesian theory[5,6], an optimal estimate results from combining information about the distribution of velocities—the prior—with evidence from sensory feedback. As uncertainty increases, when playing in fog or at dusk, the system should increasingly rely on prior knowledge. To use a bayesian strategy, the brain would need to represent the prior distribution and the level of uncertainty in the sensory feedback. Here we control the statistical variations of a new sensorimotor task and manipulate the uncertainty of the sensory feedback. We show that subjects internally represent both the statistical distribution of the task and their sensory uncertainty, combining them in a manner consistent with a performance-optimizing bayesian process[4,5]. The central nervous system therefore employs probabilistic models during sensorimotor learning.

Subjects reached to a visual target with their right index finger in a **virtual-reality set-up that allowed us to displace the visual feedback of their finger laterally relative to its actual location** (Fig. 1a; see Methods for details). On each movement, the lateral shift was randomly drawn from a prior distribution that was gaussian with a mean shift of 1 cm to the right and a standard deviation of 0.5 cm (Fig. 1b). We refer to this distribution as **the true prior**. During the movement, **visual feedback of the finger position was only provided briefly, midway through the movement**. We manipulated the reliability of this visual feedback on each trial. This feedback was either provided clearly (j0 condition, in which the uncertainty comes from intrinsic processes only), blurred to increase the uncertainty by a medium (jM) or large (jL) amount, or was withheld altogether leading to infinite uncertainty (j1). Visual information about the position of the finger at the end of the movement was provided only on clear feedback trials (j0) and subjects were instructed to get as close to the target as possible on all trials

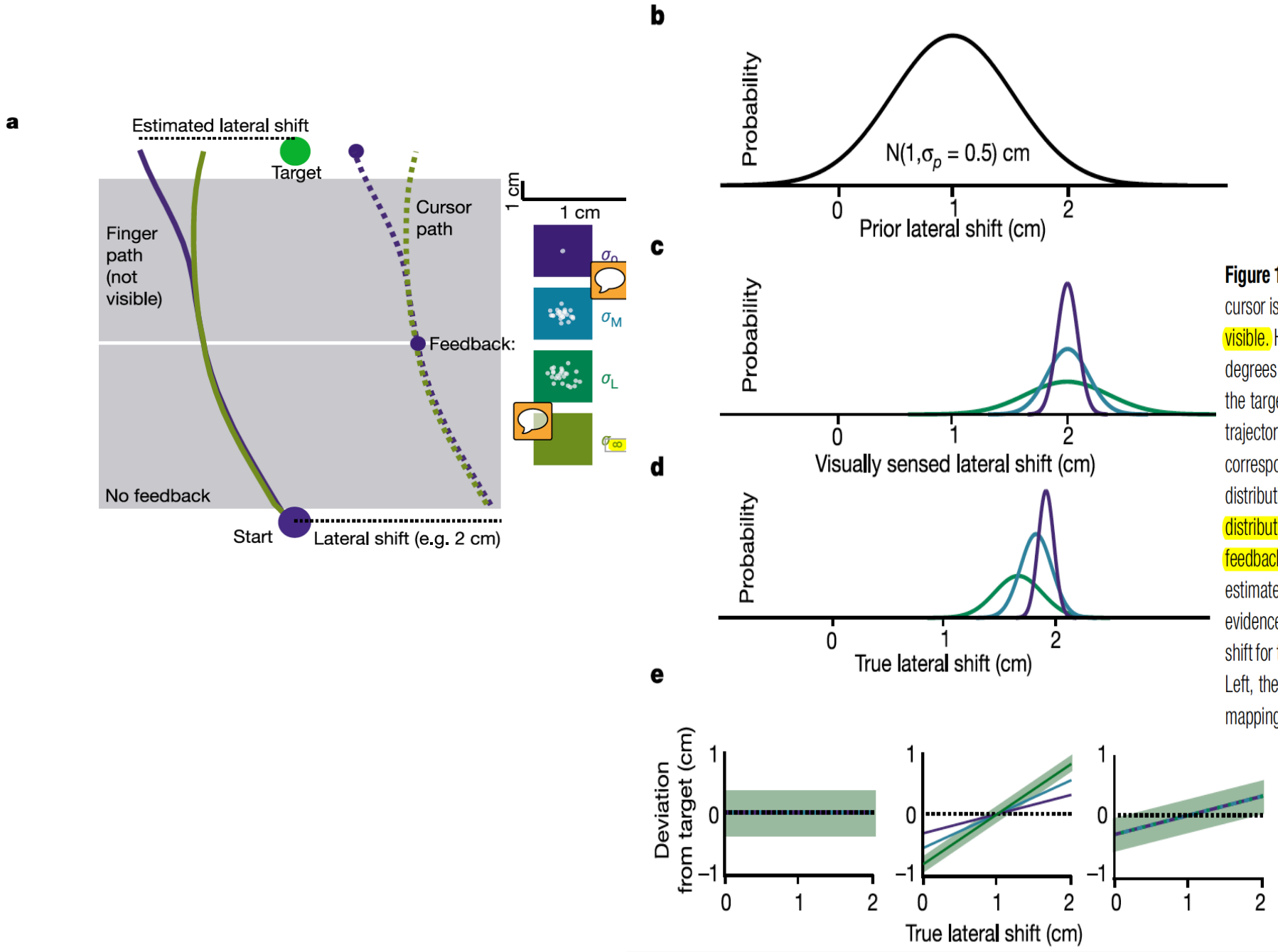


Figure 1 The experiment and models. **a**, As the finger moves from the starting circle, the cursor is extinguished and shifted laterally from the true finger location. The hand is never visible. Halfway to the target, feedback is briefly provided clearly (σ_0) or with different degrees of blur (σ_M and σ_L), or withheld (σ_∞). Subjects are required to place the cursor on the target, thereby compensating for the lateral shift. The finger paths illustrate typical trajectories at the end of the experiment when the lateral shift was 2 cm (the colours correspond to two of the feedback conditions). **b**, The experimentally imposed prior distribution of lateral shifts is gaussian with a mean of 1 cm. **c**, A diagram of the probability distribution of possible visually experienced shifts under the clear and the two blurred feedback conditions (colours as in **a**) for a trial in which the true lateral shift is 2 cm. **d**, The estimate of the lateral shift for an optimal observer that combines the prior with the evidence. **e**, The average lateral deviation from the target as a function of the true lateral shift for the models (for σ_L the green shading shows the variability of the lateral deviation). Left, the full compensation model; middle, the bayesian probabilistic model; right, the mapping model (see the text for details).

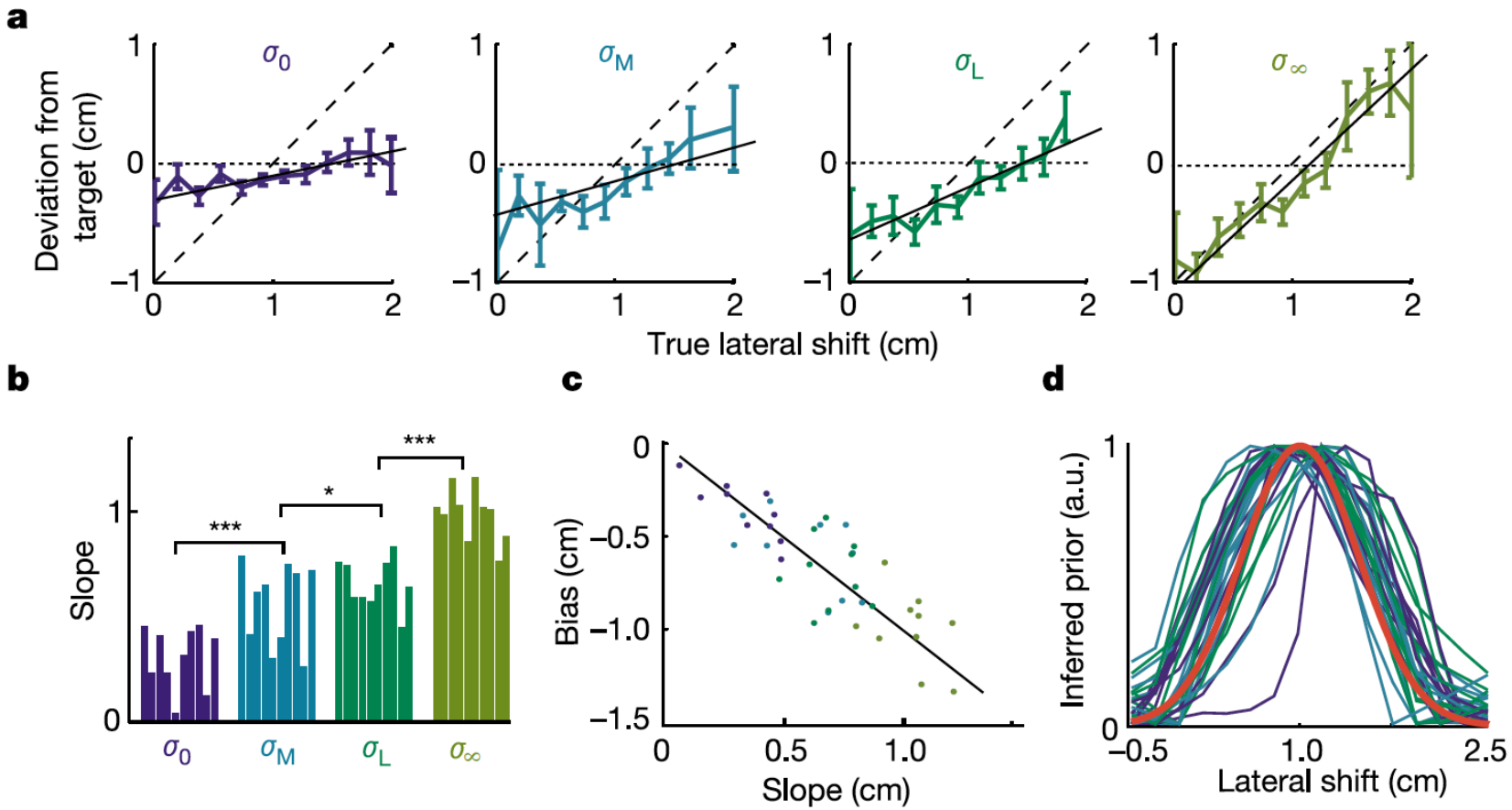


Figure 2 Results for a gaussian distribution. Colour codes as in Fig. 1. **a**, The lateral deviation of the cursor at the end of the trial as a function of the imposed lateral shift for a typical subject. Error bars denote s.e.m. The horizontal dotted lines indicate the prediction from the full compensation model and the dashed line is the fit for a model that ignores sensory feedback on the current trial and corrects only for the mean over all trials. The solid line is the bayesian model with the level of uncertainty fitted to the data. **b**, The slopes for the linear fits are shown for the full population of subjects. On the basis of the hypothesis that the slope should increase with increasing visual uncertainty, we performed a repeated-measures analysis of variance on the slope, with visual uncertainty as a factor (main effect of visual uncertainty $F_{3,27} = 82.7$; $p < 0.001$). Planned comparisons of the slopes between adjacent uncertainty levels were all significant (asterisk, $p < 0.05$; three asterisks, $p < 0.001$). **c**, The bias against gain for the linear fits for each subjects and condition. The solid line shows the bayesian solutions. **d**, The inferred priors and the true prior (red) for each subject and condition.

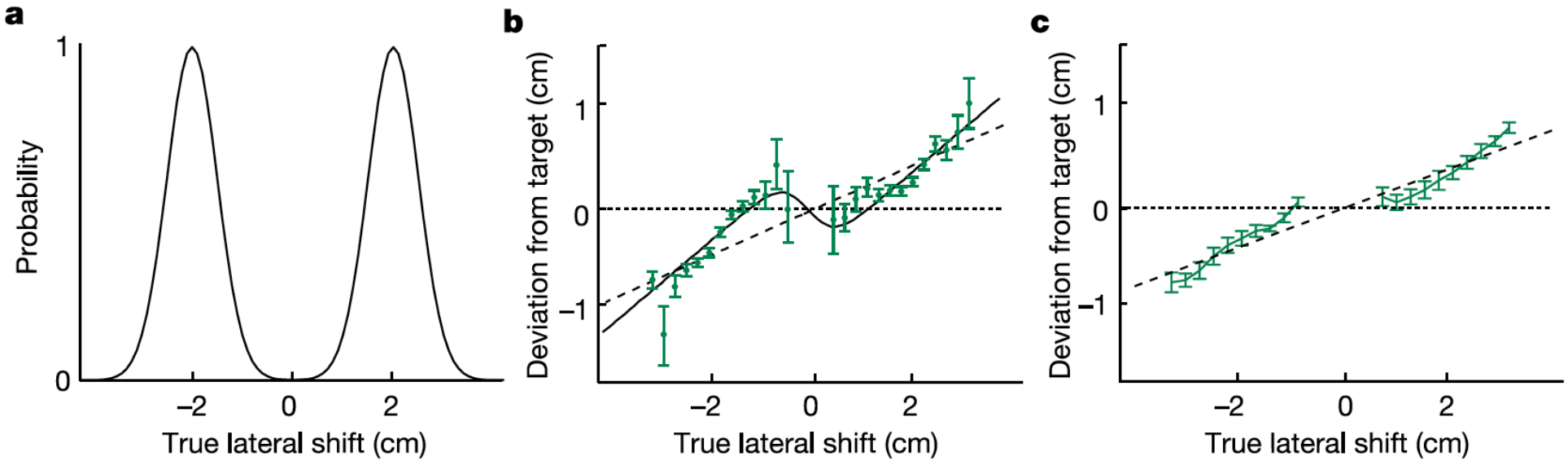


Figure 3 Results for a mixture of gaussian distributions. **a**, The experimentally imposed prior distribution of lateral shifts is a mixture of two gaussians. **b**, The lateral deviation of the cursor at the end of the trial as a function of the true lateral shift for a typical subject. Error bars denote s.e.m. The horizontal dotted lines indicate the prediction from the full compensation model, the dashed line is the fit for a bayesian model with a single gaussian prior, and the solid line is the fit for a bayesian model with a prior that is a mixture of two gaussians. **c**, The lateral deviation across subjects (mean ± s.e.m. across subjects) is shown with a linear regression fit, demonstrating the nonlinearity of the data.

Decision Theory: What "Should" the Nervous System Do?

Konrad Körding

Science **318**, 606 (2007);
DOI: 10.1126/science.1142998

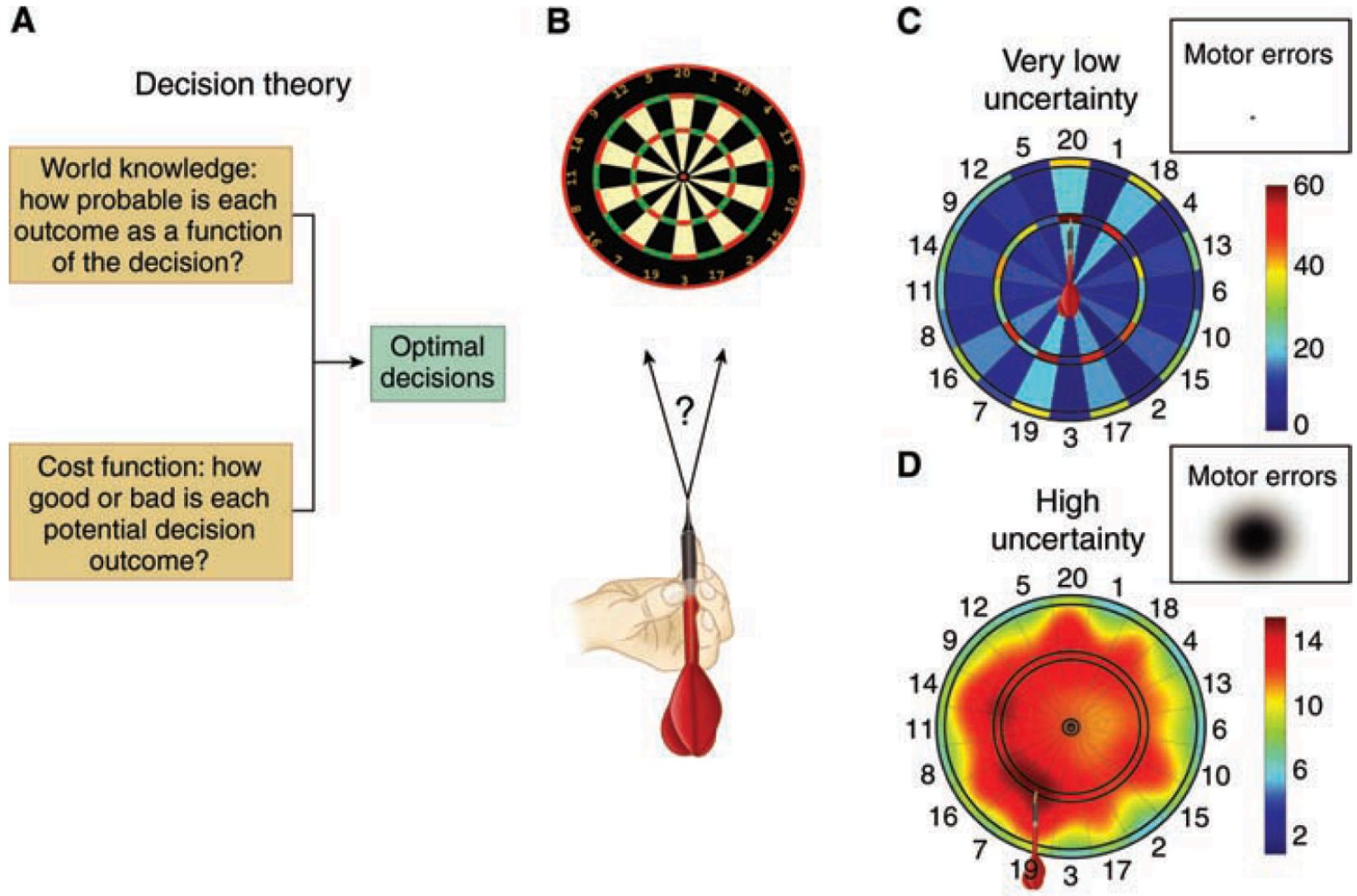


Fig. 1. Decision theory. (**A**) To make optimal decisions, we need to combine what we know about the world with our utility function measuring how good or bad potential outcomes are. (**B**) In the example of playing darts, we need to decide where to aim. (**C**) As a function of the aiming point, the expected score is shown for an unbelievably good darts player with almost no movement errors. (**D**) As in (C) but for a mediocre darts player with large motor errors.

A
utility function U(outcome) measures how good or bad any possible decision outcome is. If dart players could choose where the dart will hit the board, they would choose the position that yields the most points and would thus maximize utility.

Although we can freely make decisions, we cannot directly choose the decision outcomes. If we always aim for the same position a, say the center of the bull’s eye, and throw many darts, we will produce a distribution of dart positions, x, on the dart board (Fig. 1C, inset).Within decision theory, this probability distribution is denoted p(outcome = x|decision = a). If we aim at the position on the board that gives the highest score, we may instead hit a neighboring area of the dartboard and receive a low score. Depending on the position we aim at, different scores become more or less likely

DO NOT SHOW THIS SLIDE

Bayesian statistics defines how uncertain pieces of information may be combined into a joint estimate. New information (called a likelihood) needs to be combined or integrated with information from the past (called a prior). Similar problems occur when information from several cues, for example, proprioceptive and visual, needs to be combined into a joint estimate. Bayesian decision theory (13), the use of Bayesian statistics in a decision framework, defines how our beliefs should be combined with our utility function**. Because most if not all of our decisions are made in the presence of uncertainty**

understanding the way the nervous system deals with uncertainty is central to understanding its normal mode of operation.

DO NOT SHOW THIS SLIDE

Integration of priors and likelihoods. To calculate the probabilities of outcomes, it is often necessary to update our belief from the past (prior) with new knowledge (likelihood). For example, when we play tennis it is helpful to estimate where the ball will land.

DO NOT SHOW THIS SLIDE

Decision-Making

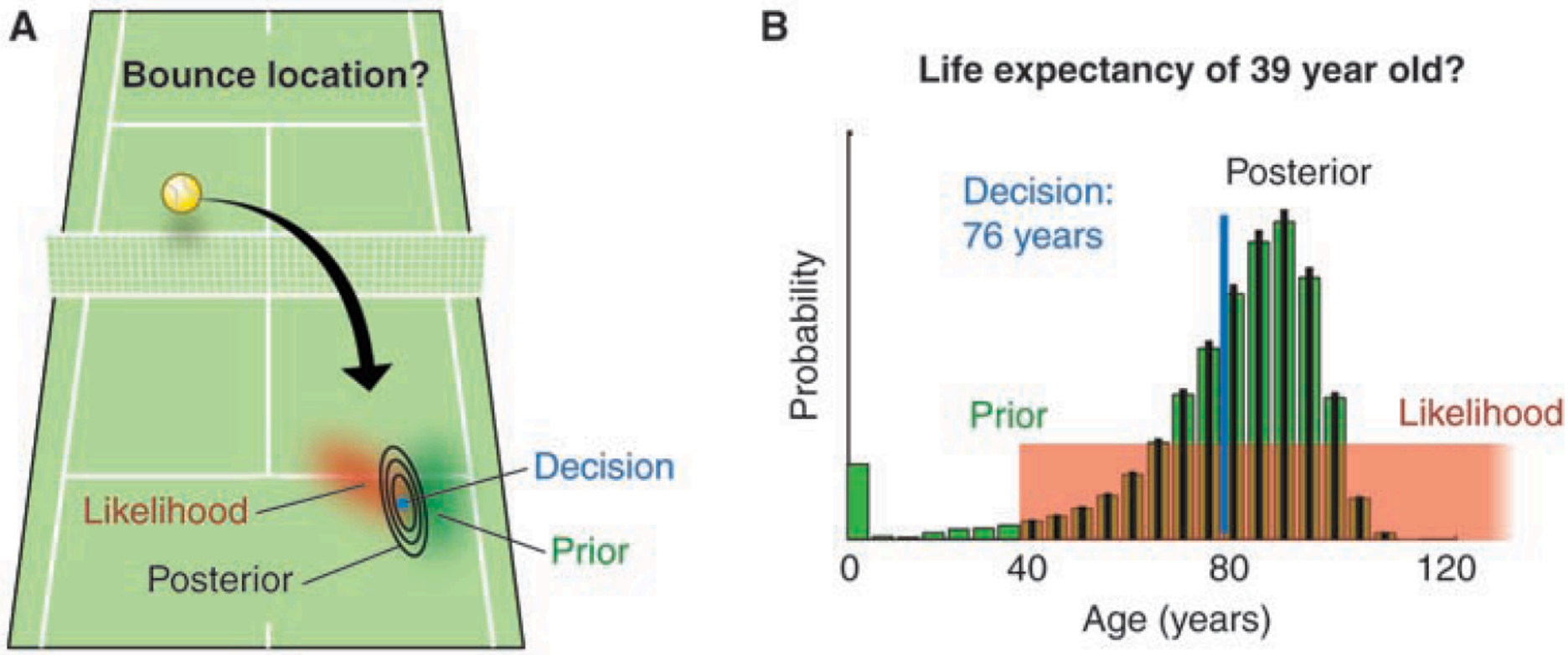


Fig. 2. (**A**) In the example of tennis, people need to combine what they know from before (prior, green) with what they currently see (likelihood, red). That way we can estimate the posterior (black contour lines) to make an optimal perceptual decision (blue). (**B**) Similarly if we estimate the life expectancy of a person who is 39 years old, we need to combine what we know from before (prior, histogram of lifetimes, green) with our new information (person survived 39 years, likelihood, red) to come up with an optimal estimate.

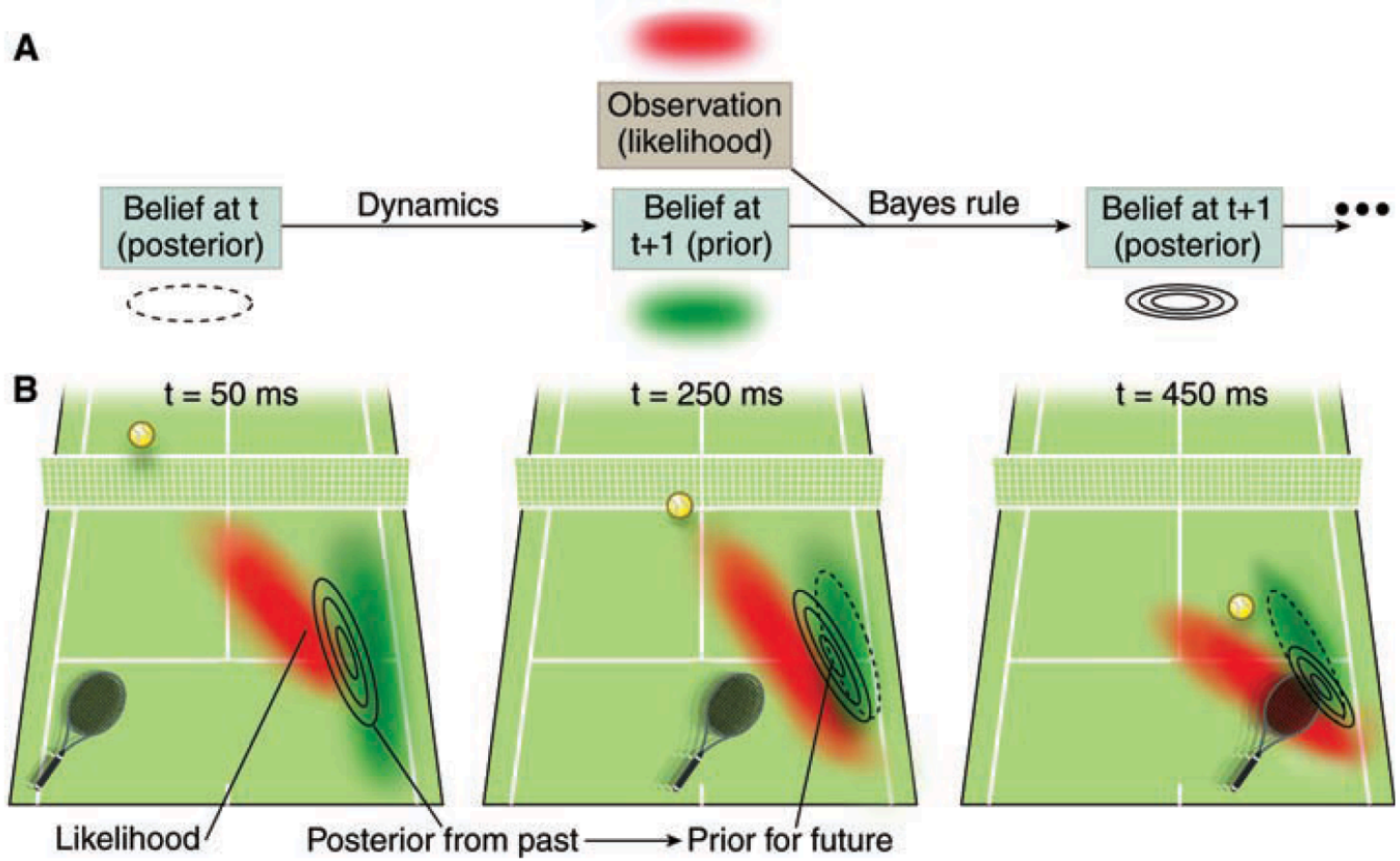


Fig. 3. Integration of information over time. (**A**) A diagram of a Kalman filter is shown. At any point of time *t*, the person has a belief about the state of the world. The person then updates this belief with a model of the dynamics of the world (e.g., gravity) to calculate the belief at the next point of time. This belief (prior) is then combined with new sensory information (likelihood) using Bayes's rule to calculate the belief at the next time step. The ellipses indicate probability distributions sketched in (B). (**B**) To estimate the position of a ball hitting the ground, people continuously update their beliefs with incoming sensory information, yielding precise estimates. The posterior of the previous time step is the prior for the new one: The dashed line indicating the previous posterior is identical to the one standard deviation line of the prior (green).

Integration of priors and likelihoods.
To calculate the probabilities of outcomes, it is often necessary to update our belief from the past (prior) with new knowledge (likelihood). For example, when we play tennis it is helpful to estimate where the ball will land. The visual system, although noisy, still provides us with an estimate or a likelihood of where the ball will land (sketched in red in Fig. 2A). This knowledge may be combined with information obtained from experience; the positions where the ball may land are not uniformly distributed over the court. The locations may be clustered near the boundary lines, where it is most difficult to return the ball. This distribution of positions is called the prior (sketched in green in Fig. 2A**). Bayes's rule states that how the probability of the ball landing at position x given our observation o (posterior) needs to be estimated as**

$$\underbrace{p(x|o)}_{\text{posterior}} = \underbrace{p(x)}_{\text{prior}} \underbrace{p(o|x)}_{\text{likelihood}} / p(o)$$

Step 2

These assumptions turn the mathematics of probability theory into **an engine of inference**, a means of weighing each of a set of mutually exclusive and exhaustive hypotheses H to determine which best explain the observed data. Probability theory tells us how to compute the degree of belief in some hypothesis hi, given some data d.

Step 1

A central assumption is that degrees of belief can be represented as probabilities: that our conviction in some hypothesis h can be expressed as a real number ranging from 0 to 1, where 0 means something like \h is completely false" and 1 that \h is completely true."

Step 3

Computing degrees of belief as probabilities depends on two components. One, called the **prior probability** and denoted **P(hi), captures how much we believe in hi prior to observing the data d**. The other, called ***the likelihood and denoted P(d/hi)***, captures the probability with which we would expect to observe the data d if hi were true. These combine to yield the **posterior probability of hi**, given via Bayes' Rule

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{\sum_{h_j \in \mathcal{H}} P(d|h_j)P(h_j)}.$$

Step 4

The denominator in Equation 1 provides a normalizing term which is the sum of
the probability of each of the possible hypotheses under consideration; this ensures that **Bayes' Rule will reflect the proportion of all of the probability that is assigned to any single hypothesis hi, and (relatedly) that the posterior probabilities of all hypotheses sum to one.**

$$\frac{}{\sum_{h_j \in \mathcal{H}} P(d|h_j)P(h_j)}$$

law of conservation of belief

This captures what we might call the "**law of conservation of belief**": a
rational learner has a fixed "mass" of belief to allocate over different hypotheses, and the act of observing data just pushes this mass around to different regions of the hypothesis space. If the data lead us to strongly believe one hypothesis, we must decrease our degree of belief in all other hypotheses. By contrast, if the data strongly disfavor all but one hypothesis, then (to paraphrase Sherlock Holmes) whichever remains, however implausible a priori, is very likely to be the truth.

Interoceptive predictions in the brain

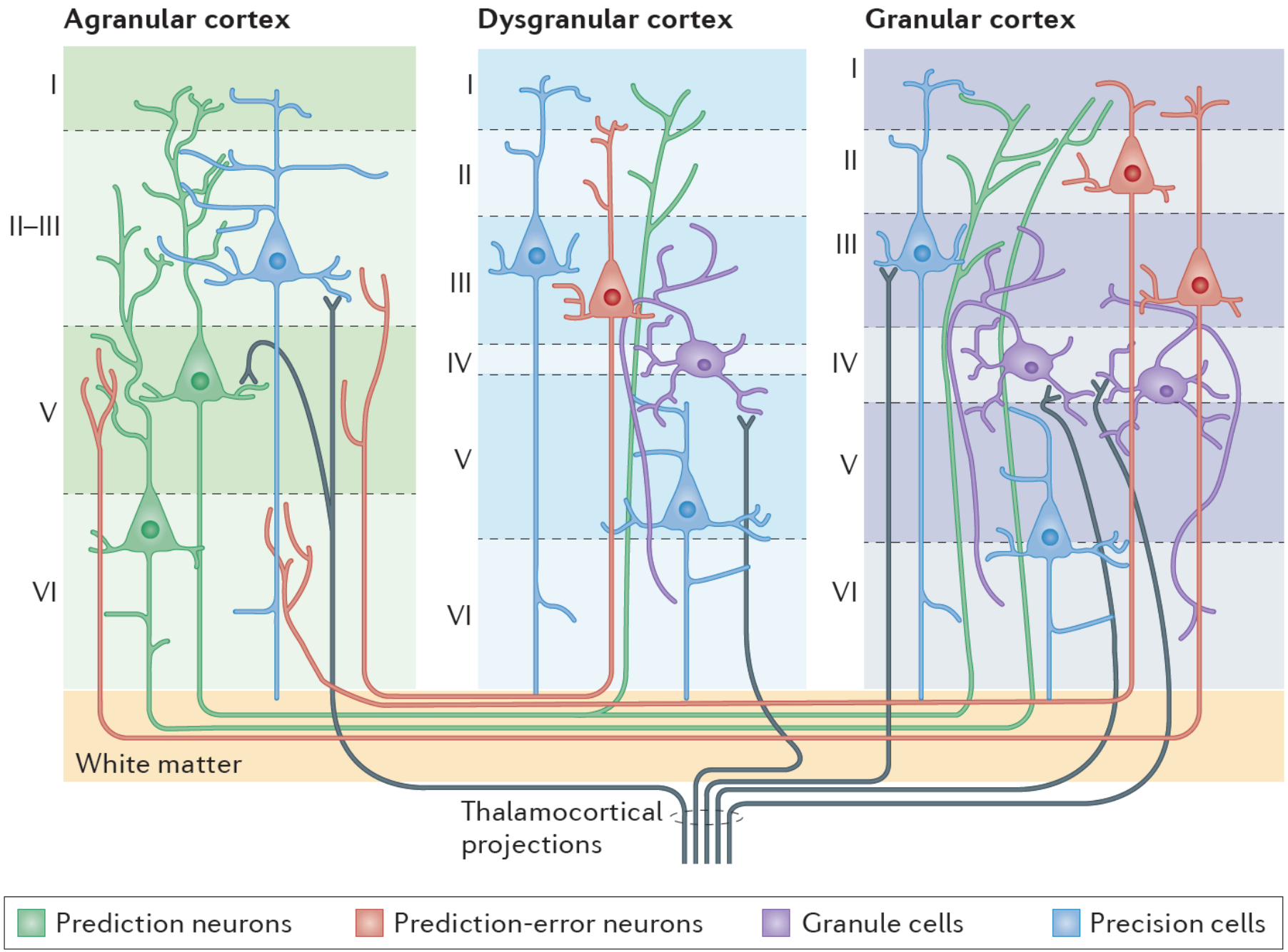
Lisa Feldman Barrett and W. Kyle Simmons

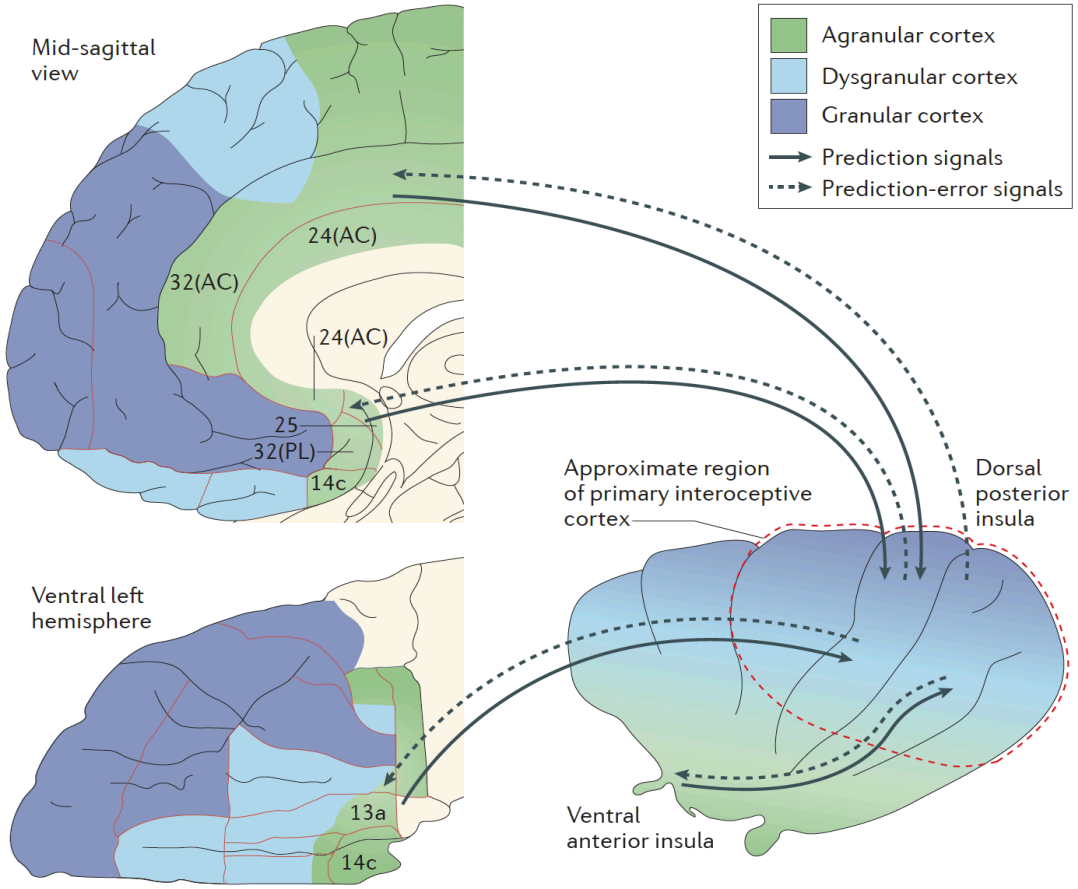
Abstract | Intuition suggests that perception follows sensation and therefore bodily feelings originate in the body. However, recent evidence goes against this logic: interoceptive experience may largely reflect limbic predictions about the expected state of the body that are constrained by ascending visceral sensations. In this Opinion article, we introduce the Embodied Predictive Interoception Coding model, which integrates an anatomical model of corticocortical connections with Bayesian active inference principles, to propose that agranular visceromotor cortices contribute to interoception by issuing interoceptive predictions. We then discuss how disruptions in interoceptive predictions could function as a common vulnerability for mental and physical illness.

According to this active inference account,
the brain forms neural representations that are constructed from previous experience. These function as a generative model of how stimuli in the environment cause sensations. Rather than neurons simply lying dormant until information arrives via the external sensors
the body (that is, the eyes, ears and taste receptors, among others), the brain anticipates incoming sensory inputs, which it implements as predictions that cascade throughout the cortex. As predictions propagate across cortical regions –following their roughly **centrifugal** connections–

they modulate the firing of neurons within cortical columns in anticipation of these regions receiving actual sensory sensation –that is, the "prediction error"

In this active inference framework,
perception and action are tightly coupled, with both arising from the brain's hypotheses about the world and constrained by sensory inputs from the world. By this account, **action drives perception to reduce prediction error.**


Agranular cortex
Dysgranular cortex
Granular cortex
I
II–III
V
VI
I
II
III
IV
V
VI
I
II
III
IV
V
VI
White matter
Thalamocortical projections
Prediction neurons
Prediction-error neurons
Granule cells
Precision cells



According to this model, we suggest that **agranular visceromotor cortices** — including the cingulate cortex (Brodmann area 24 (BA24), BA25 and BA32), the posterior ventral medial prefrontal cortex (BA14c), the posterior orbitofrontal cortex (BA13a) and the most ventral portions of the anterior insula — **estimate the balance between the autonomic, metabolic and immunological resources that are available to the body, and the predicted requirements of the body, based on past experience.**

visceromotor cortices simultaneously issue predictions of the interoceptive signals that are expected to arise as consequences of those allostatic visceral changes to the primary interoceptive sensory cortex (see the figure)

The granular cortex in primary interoceptive sensory regions of the mid- and posterior insula are architecturally well suited for computing and transmitting prediction error and for propagating prediction-error signals back to visceromotor regions to modify predictions.

This means that **interoceptive perception is largely a construction of beliefs that are kept in check by the actual state of the body** (rather than vice versa). What you experience is in large part a reflection of what your brain predicts is going on inside your body, based on past experience.

Many of the key regions that have been implicated in the pathophysiology of depression, such as the subgenual and subcallosal ACC and the anterior insula are agranular visceromotor limbic regions within the interoceptive system that is proposed here. It is well known that structural abnormalities and chronically hyperactive metabolism within agranular visceromotor regions precede the onset of depression (for example, see REFS **93,94**). According to the EPIC model, the predictions that result from these structurally or functionally abnormal visceromotor regions may result in a break-down of body systems that are needed to maintain homeostasis in response to stressors or everyday events that are perceived as stressors

Drevets, W. C. *et al.* Subgenual prefrontal cortex abnormalities in mood disorders. *Nature* **386**, 824–827 (1997).
Boes, A. D., McCormick, L. M., Coryell, W. H. & Nopoulos, P. Rostral anterior cingulate cortex volume correlates with depressed mood in normal healthy children. *Biol. Psychiatry* **63**, 391–397 (2008).
Levesque, M. L. *et al.* Altered patterns of brain activity during transient sadness in children at familial risk for major depression. *J. Affect. Disord.* **135**, 410–413 (2011).

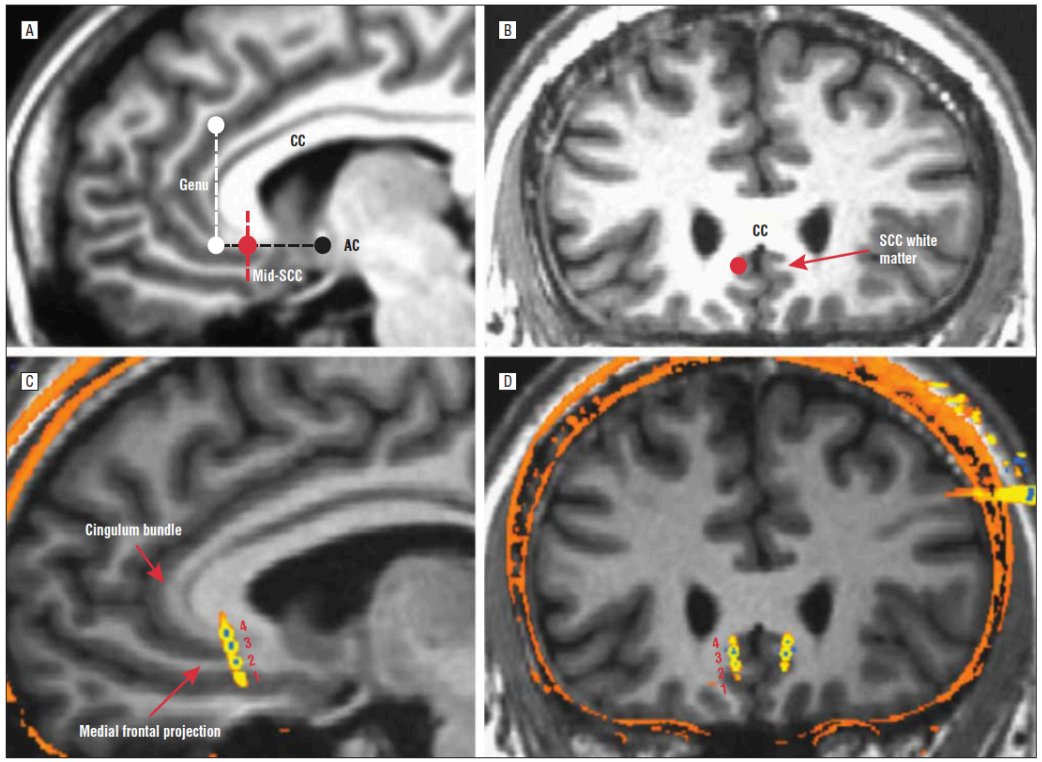
In the long term, this chronic imbalance — which is caused by constantly predicting the need for more metabolic energy to meet the demands of stressors[95] — can produce the well-known depression-related disruption and eventual downregulation of hypothalamus–pituitary–adrenal (HPA)-axis negative-feedback loops, resulting in chronic hypercortisolaemia (*Mol. Psychiatry* **7**:254-275 (2002)). This in turn can promote a pro-inflammatory state that is associated with increased levels of cytokines and activated immune biochemical pathways (*Nat. Rev. Neurosci. 9:46-56 (2008)).*

Gold, P. W. & Chrousos, G. P. Organization of the stress system and its dysregulation in melancholic and atypical depression: high versus low CRH/NE states. *Mol. Psychiatry* **7**, 254–275 (2002).

Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W. & Kelley, K. W. From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat. Rev. Neurosci.* **9**, 46–56 (2008).

To **reduce prediction error**, limbic visceromotor cortices begin guiding the body towards a constellation of sickness behaviors associated with fatigue and negative affect that are designed to reduce activity and energy expenditure. Collectively, these behaviors would be the initial behavioral symptoms of depression

EPIC model may also inform treatment of depression and some anxiety disorders. For example, deep brain stimulation of the connections that project out of the subcallosal cingulate cortex in a region of the visceromotor system (particularly **in BA25**) is effective in remitting treatment-resistant depression (Kennedy et al 2011; Riva-Posse et al 2014)



Kennedy, S. H. *et al.* Deep brain stimulation for treatment-resistant depression: follow-up after 3 to 6 years. *Am. J. Psychiatry* **168**, 502–510 (2011).
Riva-Posse, P. *et al.* Defining critical white matter pathways mediating successful subcallosal cingulate deep brain stimulation for treatment-resistant depression. *Biol. Psychiatry* **76**, 963–969 (2014).

Perhaps our most speculative but innovative hypothesis concerns the relationship between interoceptive predictions and certain physical illnesses that often co-occur with depression, such as diabetes, heart disease and cancer[115]. ***Aberrant interoceptive predictions and the compounding allostatic consequences that may result could help to explain the links among these disorders***. For example, many of the same regions within the interoceptive system that show morphological changes in psychiatric illness[90] and chronic pain[116] also show morphological changes with accumulated stress across the lifespan[117] and leave individuals more vulnerable to these metabolic illnesses and with increased risk of mortality[118], particularly if this stress occurred in childhood[119]. All of these illnesses are also linked to homeostatic and inflammatory mechanisms.

Moussavi, S. *et al.* Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858 (2007).

Goodkind, M. *et al.* Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305–315 (2015).

Piazza, J. R., Charles, S. T., Sliwinski, M. J., Mogle, J. & Almeida, D. M. Affective reactivity to daily stressors and long-term risk of reporting a chronic physical health condition. *Ann. Behav. Med.* **45**, 110–120 (2013).

. Dannlowski, U. *et al.* Limbic scars: long-term consequences of childhood maltreatment revealed by functional and structural magnetic resonance imaging. *Biol. Psychiatry* **71**, 286–293 (2012).

. Smallwood, R. F. *et al.* Structural brain anomalies and chronic pain: a quantitative meta-analysis of gray matter volume. *J. Pain* **14**, 663–675 (2013).

. Ganzel, B. L., Morris, P. A. & Wethington, E. Allostasis and the human brain: integrating models of stress from the social and life sciences. *Psychol. Rev.* **117**, 134–174 (2010).

In a Bayesian sense, the effects of CBT may reflect changes in the way that precision-weighting pyramidal cells in the viscerosensory cortex adjust the weight of prediction-error signals that are communicated to agranular cortices, thus altering the sampling of inputs that become the 'empirical priors' in subsequent predictions. Interestingly, emerging evidence indicates that the **activity within agranular visceromotor cortices predicts whether CBT or pharmacotherapy will be more effective as a treatment option**

McGrath, C. L. *et al.* Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry* **70**, 821–829 (2013).
McGrath, C. L. *et al.* Pretreatment brain states identify likely nonresponse to standard treatments for depression. *Biol. Psychiatry* **76**, 527–535 (2014).

COGNITIVE DEVELOPMENT

Observing the unexpected enhances infants' learning and exploration

Aimee E. Stahl* and Lisa Feigenson

Stahl and Feigenson (2015) Science 348:91-94

The Bayesian Brain and Learning
Stahl and Feigenson (2015) Science 348:91-94

Given the overwhelming quantity of information available from the environment, how do young learners know what to learn about and what to ignore? 11-month-old infants use ***violations of prior expectations as special opportunities for learning***. The infants were shown events that violated expectations about object behavior or events that were nearly identical but did not violate expectations. The sight of an object that violated expectations enhanced learning and promoted information-seeking behaviors; specifically, infants learned more effectively about objects that committed violations, explored those objects more, and engaged in hypothesis-testing behaviors that reflected the particular kind of violation seen. Thus, early in life, expectancy violations offer a wedge into the problem of what to learn.

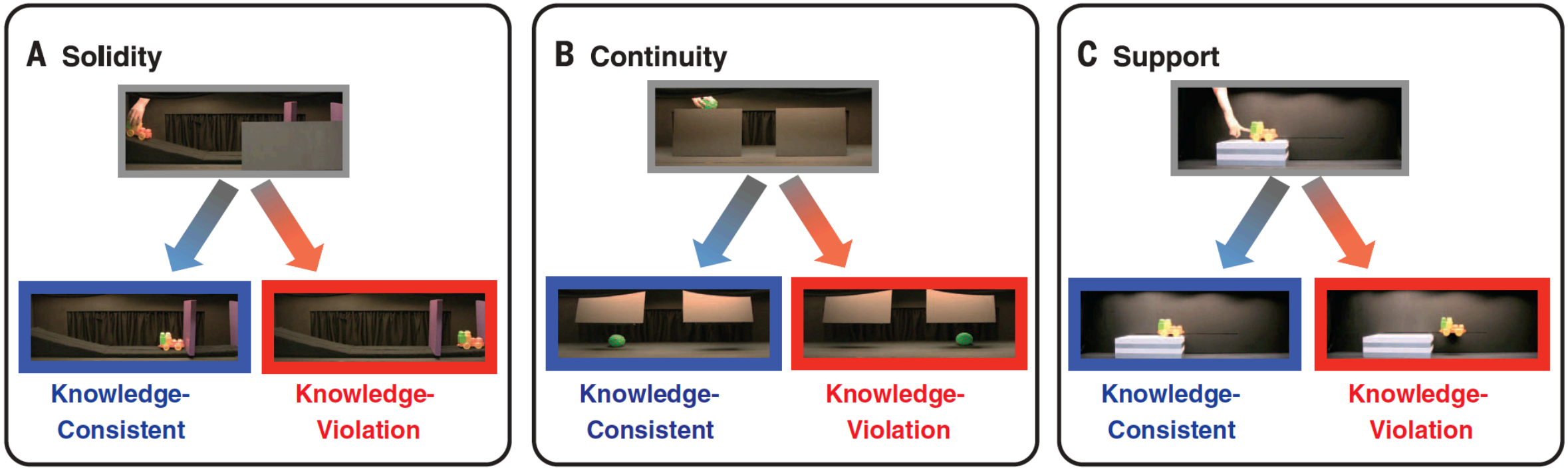


Fig. 1. Knowledge-Consistent and Knowledge-Violation outcomes in experiments 1 to 4. (**A**) Solidity events (movies S1 and S2). (**B**) Continuity events (movies S3 and S4). (**C**) Support events (movies S5 and S6).

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s2.mov

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s1.mov

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s3.mov

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s4.mov

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s5.mov

http://www.sciencemag.org/content/suppl/2015/04/01/348.6230.91.DC1/aaa3799s6.mov

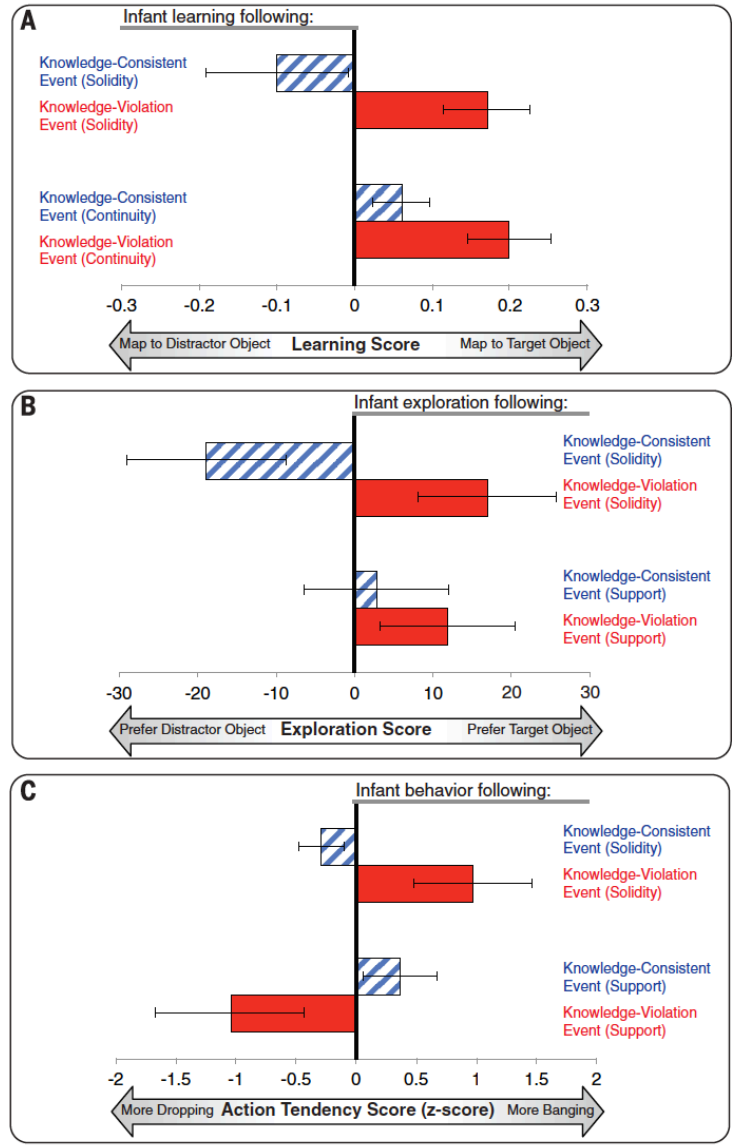


Fig. 2. Results from experiments 1 and 4. (**A**) Infants' learning after Knowledge-Consistent and Knowledge-Violation events in experiment 1. Bars represent average learning scores (proportion of looking at target object during mapping test minus proportion of looking at target object during baseline). (**B**) Infants' exploration after Knowledge-Consistent and Knowledge-Violation events in experiment 4. Bars represent looking at and/or touching the target object minus looking at and/or touching the new distractor object. (**C**) Infants' exploratory behaviors on the target object after Knowledge-Consistent and Knowledge-Violation events in experiment 4. Bars represent infants' *z*-scored object-banging behaviors minus *z*-scored object-dropping behaviors. All error bars represent SEM.

Some of
these sophisticated behaviors have been interpreted in terms of Bayesian inferences that generate knowledge by weighing new evidence against prior beliefs (29, 30). Our findings accord well with such a framework and suggest avenues to explore how violations detected in different domains of prior knowledge, or using different kinds of new evidence, shape exploration and learning throughout the life span and across species.

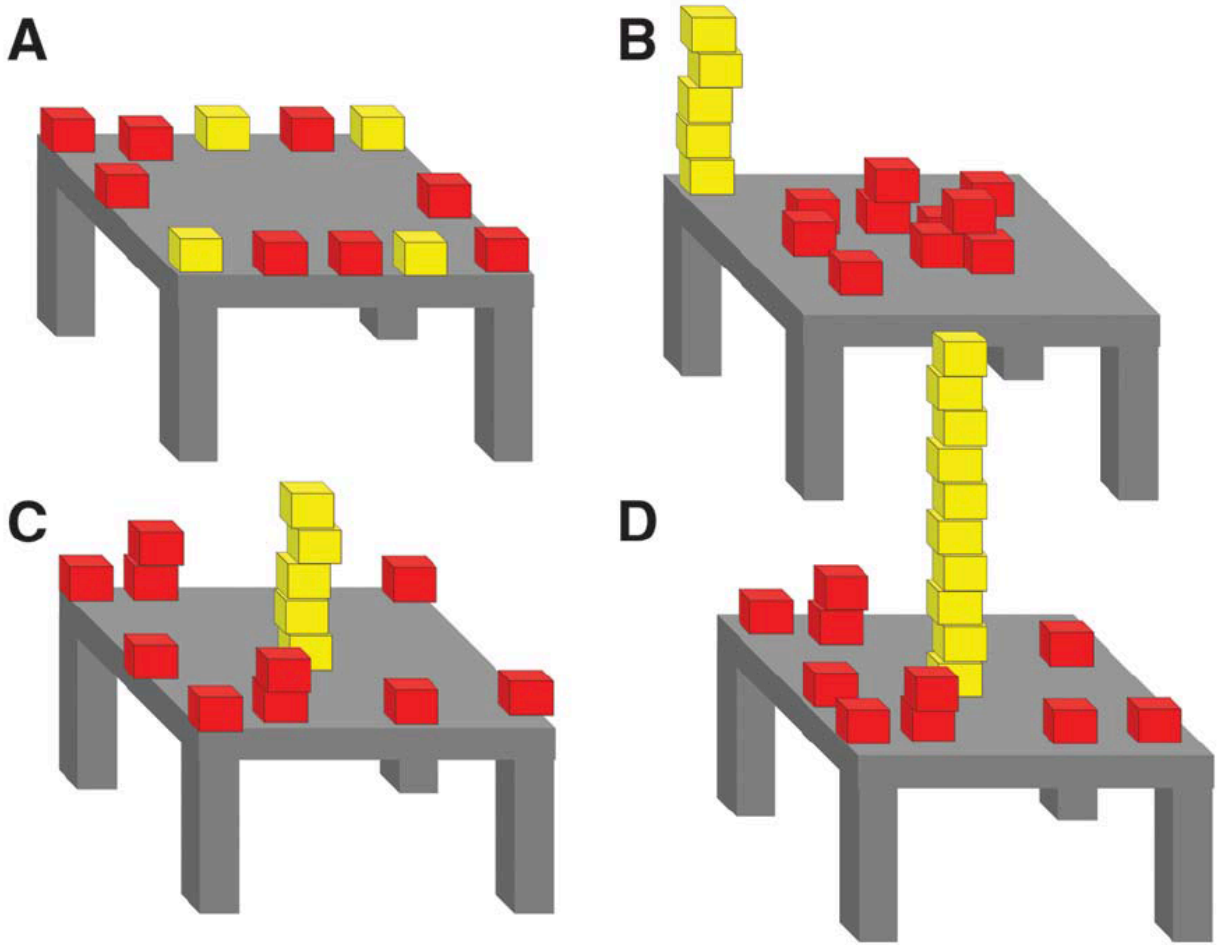
29. J. B. Tenenbaum, T. L. Griffiths, C. Kemp, Trends Cognit. Sci. 10, 309–318 (2006).
30. L. Schulz, Trends Cognit. Sci. 16, 382–389 (2012).

Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference

Ernő Téglás, Edward Vul, Vittorio Girotto, Michel Gonzalez, Joshua B. Tenenbaum, Luca L. Bonatti
Science 332:1054-1059 (2011)

Many organisms can predict future events from the statistics of past experience, but humans also excel at making predictions by pure reasoning: integrating multiple sources of information, guided by abstract knowledge, to form rational expectations about novel situations, never directly experienced. Here, we show that this reasoning is surprisingly rich, powerful, and coherent even in preverbal infants. When 12-month-old infants view complex displays of multiple moving objects, they form time-varying expectations about future events that are a systematic and rational function of several stimulus variables. ***Infants' looking times are consistent with a Bayesian ideal observer embodying abstract principles of object motion.*** The model explains infants' statistical expectations and classic qualitative findings about object cognition in younger babies, not originally viewed as probabilistic inferences.

Fig. 1. Examples of common-sense predictions based on pure reasoning. If the table in this scene is bumped so that one block falls off the table onto the floor, is it more likely to be a red or a yellow block? Intuitions will vary according to the number of blocks of each type (**A**), their arrangement into more- or less-precarious stacks and their locations on the table (**B**), and interactions between all these factors (**C** and **D**).



Science 332:1054-1059 (2011)

Twelve kinds of movies were generated by manipulating three factors relevant to predicting these outcomes: the number of objects of each type in the scene (three instances of one type and one of the other type), their physical arrangement (objects of one type were always closer to the exit before occlusion than objects of the other type), and the duration of occlusion (0, 1, or 2 s). Forming correct expectations here requires the ability to integrate these three information sources, guided by abstract knowledge about how objects move: at a minimum, qualitative knowledge about solidity (objects are unlikely to pass through walls) and spatiotemporal continuity (objects tend to move short distances over brief time intervals). Infants appear to be sensitive to each of these information sources and knowledge systems individually

http://www.sciencemag.org/content/suppl/2011/05/25/332.6033.1054.DC1/1196404s1.mov

http://www.sciencemag.org/content/suppl/2011/05/25/332.6033.1054.DC1/1196404s2.mov

http://www.sciencemag.org/content/suppl/2011/05/25/332.6033.1054.DC1/1196404s3.mov

http://www.sciencemag.org/content/suppl/2011/05/25/332.6033.1054.DC1/1196404s4.mov

http://www.sciencemag.org/content/suppl/2011/05/25/332.6033.1054.DC1/1196404s5.mov

Infants' reasoning
abilities are typically studied by measuring their looking times to visually presented events as an index of surprise: Longer looking indicates greater violation of infants' expectations relative to their prior knowledge or greater novelty relative to their interpretation of habituation stimuli. Looking time studies suggest that preverbal infants can reason about novel events depending on certain physical outcome

DO NOT INCLUDE THIS SLIDE

Looking time
studies suggest that preverbal infants can reason about novel events depending on certain physical outcomes (11, 12); object numerosities (13); other agents' beliefs, goals and behaviors (14–16); and the likely outcomes of simple random processes (17, 18).

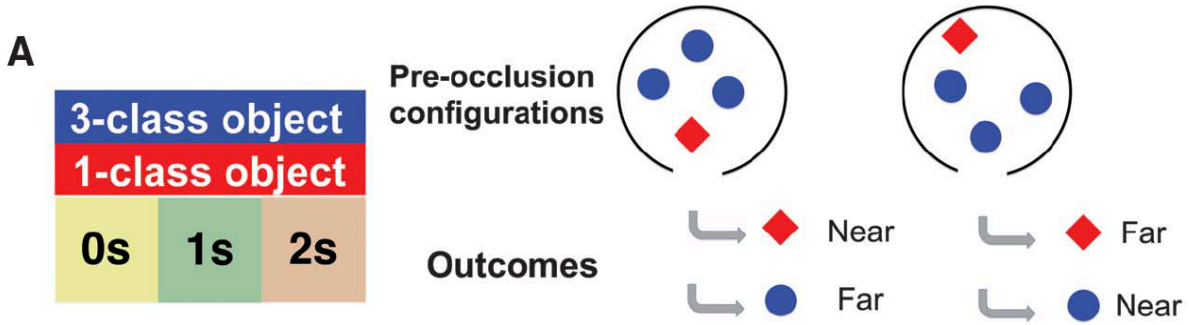
DO NOT INCLUDE THIS SLIDE

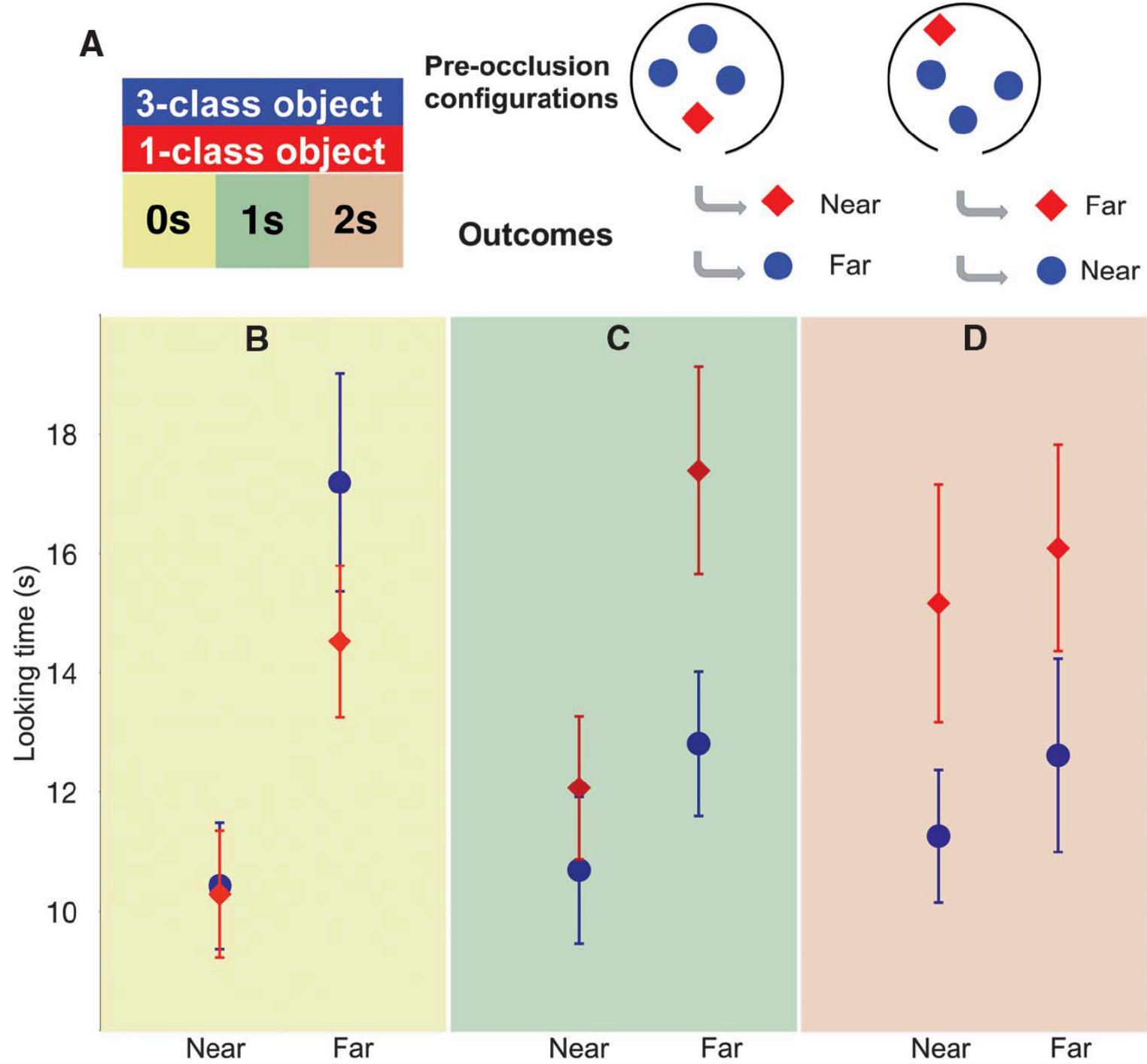
DO NOT INCLUDE THIS SLIDE

kept test events fixed and equal in salience so that infants' looking times had the potential to Show variations in degrees of belief (or conversely, degrees of surprise) as their expectations changed.

We describe a Bayesian ideal observer model that predicts infants' looking times in our studies and extends to other aspects of infants' reasoning about the physical world, giving a unifying explanation of several classic results in infant cognition.
This model shows *how powerful pure reasoning capacities could derive from the operation of probabilistic inference mechanisms constrained by abstract principles of how objects act and interact over time.*

DO NOT INCLUDE THIS SLIDE


A
3-class object
1-class object
0s
1s
2s
Pre-occlusion configurations
Outcomes
Near
Far
Far
Near


A
3-class object
1-class object
0s
1s
2s
Pre-occlusion configurations
Outcomes
Near
Far
Far
Near
B
C
D
Looking time (s)
18
16
14
12
10
Near
Far
Near
Far
Near
Far

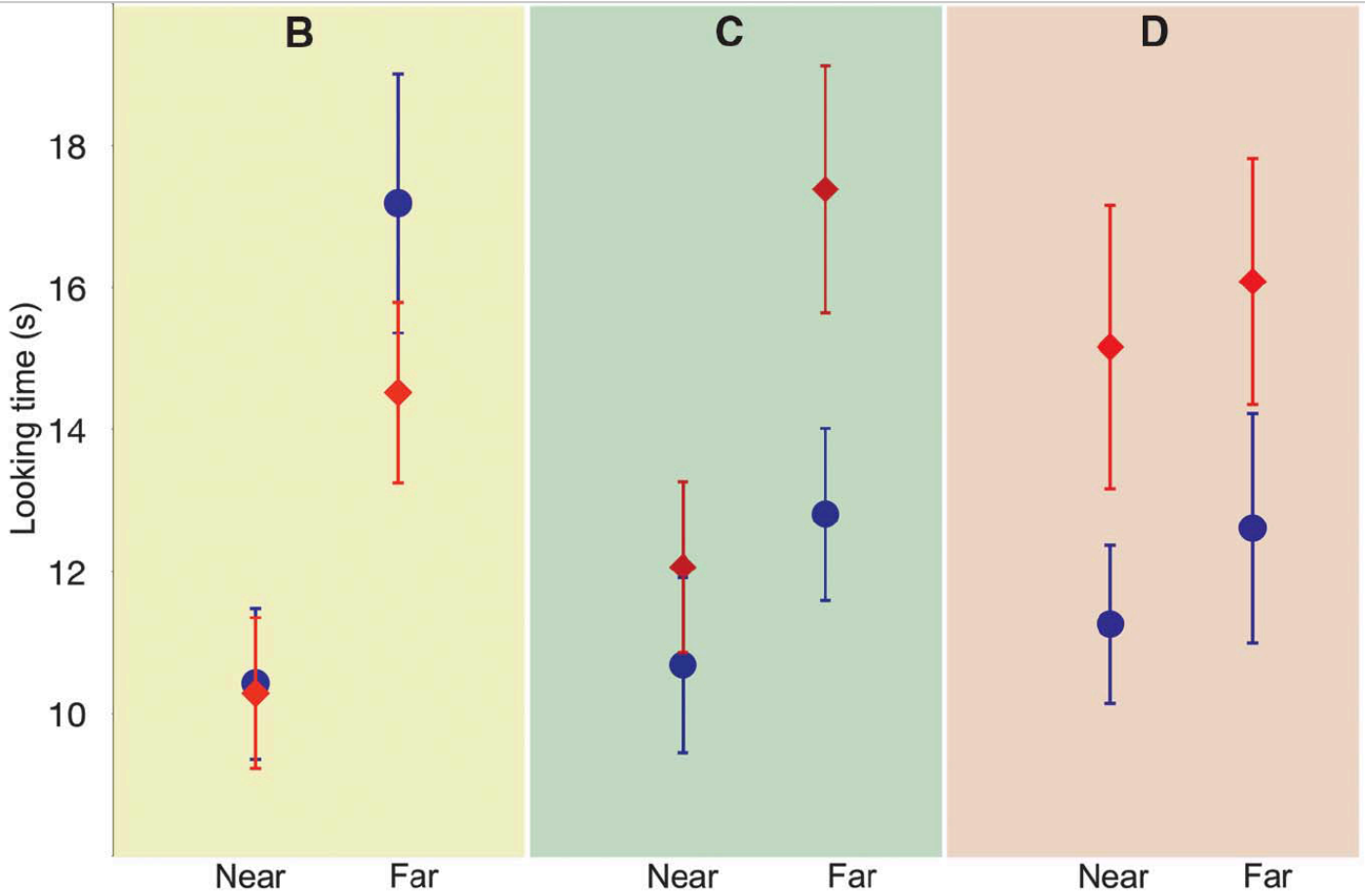
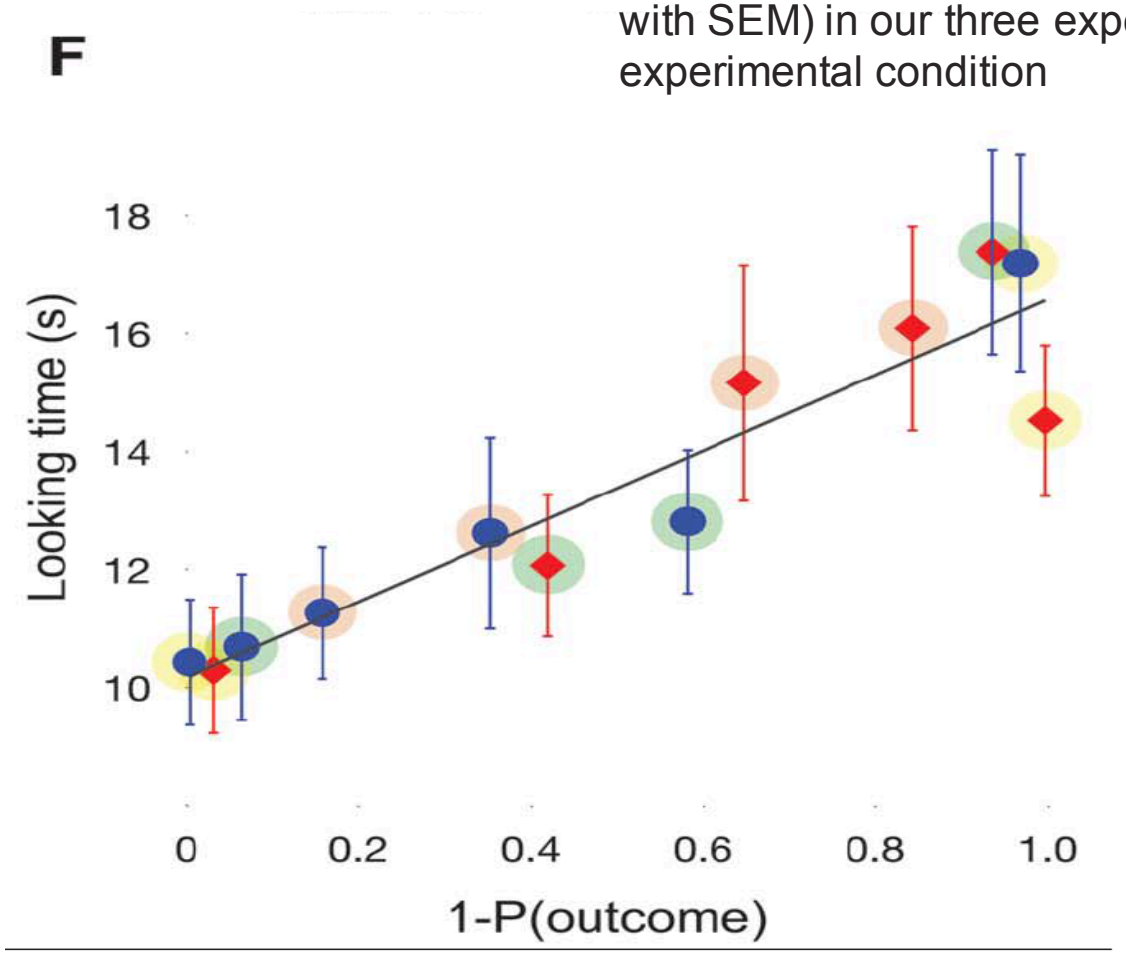


Fig. 2. Experiments probing infants' expectations in dynamic physical scenes. (**A**) Infants saw three objects of one type and one object of another type bouncing randomly inside a container. After some time, an occluder masked the objects, and one of four outcomes occurred: An object exited the container through the bottom opening that was either the common object kind or the unique object, with a position before occlusion that was either far from or near to the exit. The graph reports mean looking time (s, with SEM) of three experiments varying the duration of occlusion before the outcome. (**B**) After a short (0.04 s) occlusion, infants considered only the physical distance in forming their expectations, disregarding the number of objects of each type. (**C**) When occlusion duration was increased to 1 s, infants' looking times reflected both the number of objects of each type and their distance from the exit. (**D**) When the occlusion was longer still (2 s), infants' looking times reflected only the numerosities of each object type, regardless of their preocclusion distance from the exit.

Twelve kinds of movies were generated by manipulating three factors relevant to predicting these outcomes: the number of objects of each type in the scene (three instances of one type and one of the other type), their physical arrangement (objects of one type were always closer to the exit before occlusion than objects of the other type), and the duration of occlusion (0, 1, or 2 s). Forming correct expectations here requires the ability to integrate these three information sources, guided by abstract knowledge about how objects move: at a minimum, qualitative knowledge about solidity (objects are unlikely to pass through walls) and spatiotemporal continuity (objects tend to move short distances over brief time intervals). Infants appear to be sensitive to each of these information sources and knowledge systems individually

Correlation between the model predictions (x axis) and infant looking times (y axis, s
with SEM) in our three experiments. Each data point corresponds to one experimental condition



Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia

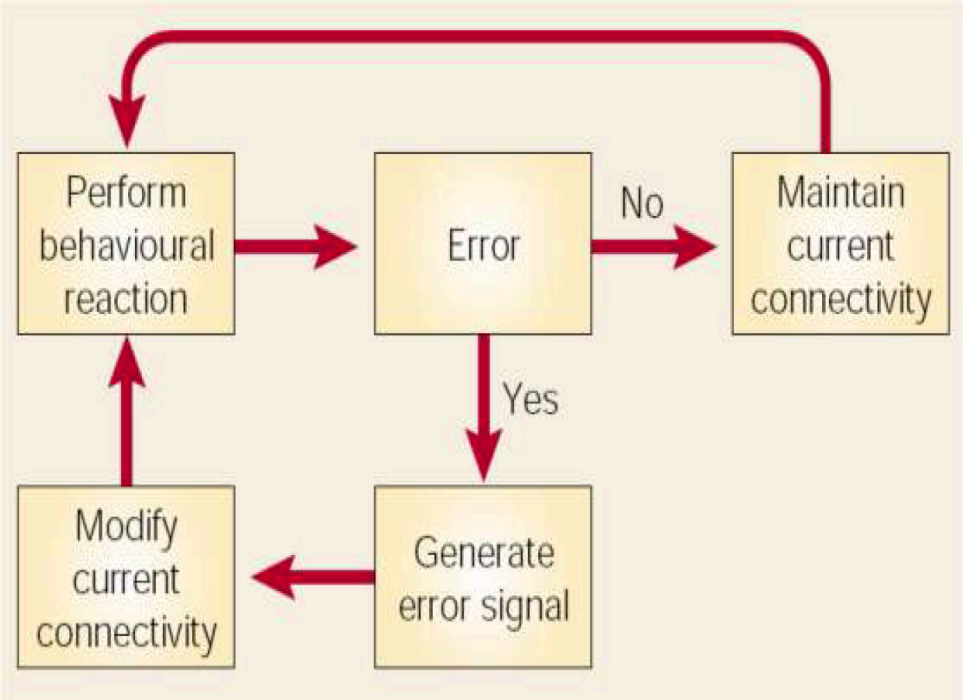
Paul C. Fletcher and Chris D. Frith‡§*

Fletcher and Frith (2009), Nature Reviews Neuroscience 10: 48-57

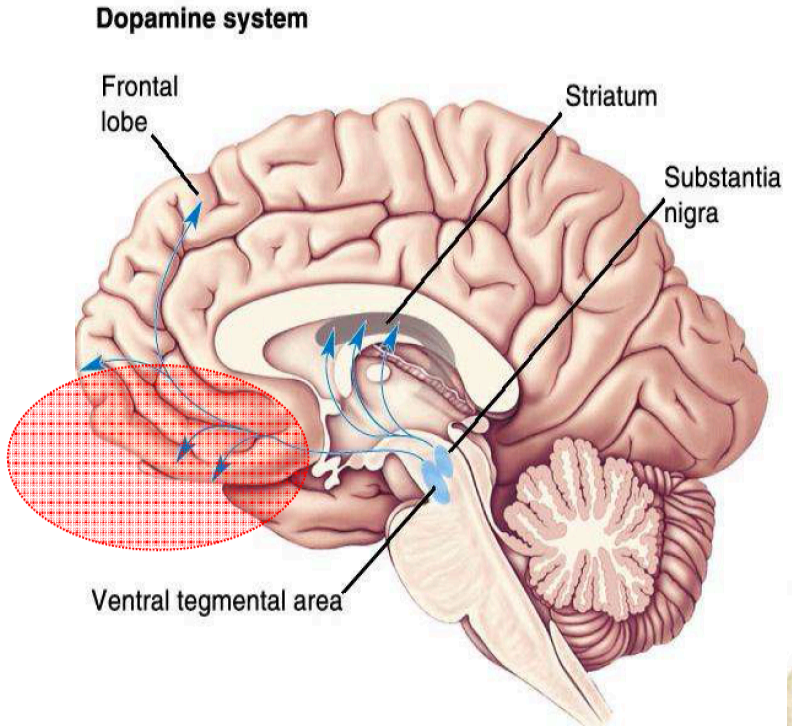
Recent advances in computational neuroscience have led us to consider the unusual perceptual experiences of patients and their sometimes bizarre beliefs as part of the same core abnormality — a **disturbance in error-dependent updating of inferences and beliefs about the world.** We suggest that it is possible to understand these symptoms in terms of a ***disturbed hierarchical Bayesian framework, without recourse to separate considerations of experience and belief.***

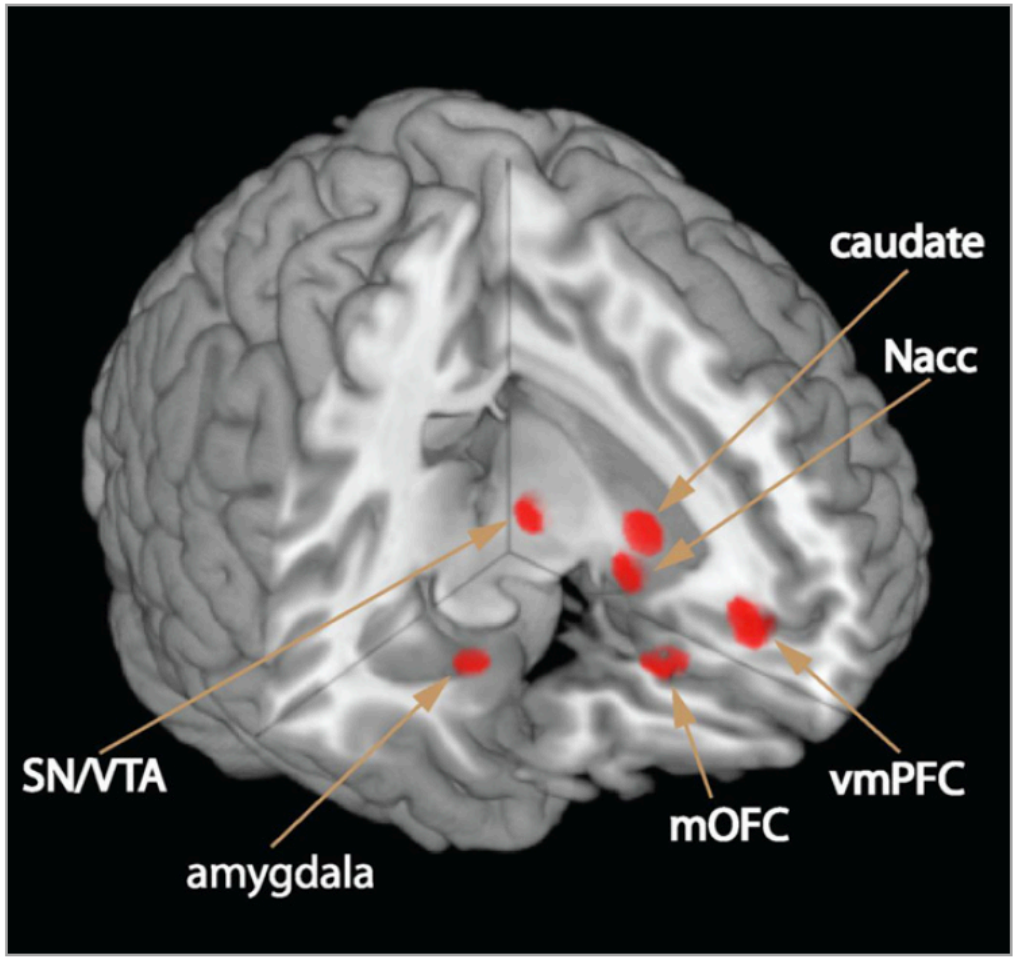
Prediction error – the discrepancy between an actually received reward and its prediction.
Learning is proportional to the prediction error.

Dopamine response = Reward occurred – Reward predicted



NATURE REVIEWS | NEUROSCIENCE
VOLUME 1 | DECEMBER 2000





Schematic illustration of the **valuation network**. Regions commonly implicated in evaluating rewards and risks in neuroeconomic imaging studies include dopaminergic neurons in the brainstem, such as substantia nigra (SN) and ventral tegmental area (VTA), which send projections to specific areas in the ***ventral striatum***, such as the caudate nucleus and nucleus ***accumbens (Nacc).*** Dopaminergic projections also modulate neuronal activity in ***ventromedial Prefrontal Cortex (vmPFC)*** and ***medial orbitofrontal cortex (vmOFC***), which have repeatedly been shown to represent reward value.

It seems that people spend most of their time with the delusion that they have an accurate representation of the world. Actually, evidence suggests that we are all rather poor at letting our sensory experience update our beliefs, and that we are susceptible to prior beliefs and social constraints that greatly limit our ability to deal with evidence rationally. For most of us, this may be manifest as poor performance when we try to deal with probabilities or as vulnerability to biases as we try to model the world. For the most part, people do not depart from the beliefs of the herd.

dopamine-neuron
firing encodes the precision or uncertainty of prediction errors and this precision weights the influence of prediction errors on inference. This is crucial for optimizing the balance between top-down prior beliefs and bottom-up sensory evidence. We can therefore speculate on the impact of abnormal dopamine-mediated neuromodulation on prediction errors. It is not the prediction errors per se that are faulty, it is the way that they are used and quantified. The size of the prediction error is meaningless without an estimate of its precision.

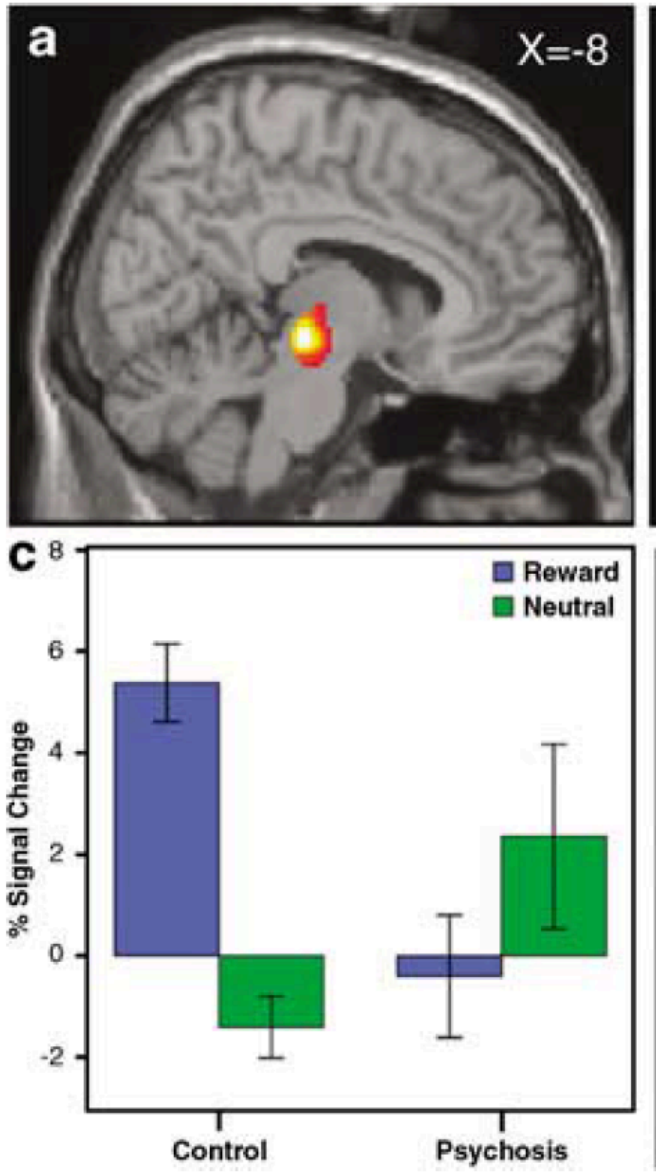
Positive Symptoms of schizophrenia may be explained by a **disruption in the prediction-error signal.** In this section, we consider more closely the possible nature of this disruption, in terms of how it might arise from abnormal dopamine neurotransmission and how this single disruption could be reflected in impairments, comparable in nature but perhaps different in expression, as we move from low-level sensory to higher-level inferential processing.

A **noisy prediction-error signal** could therefore lead to patients' strange experiences, together with their readiness to accept incidental stimuli and events as important and meaningful and to link them in unusual ways***. Persistence of the disruption up the hierarchy can mean that the attempts at the lower levels to explain the world will fail***. Achieving a world model that is not continually being signaled as wrong will require more complex changes. ***The world will feel strange, and there may be a sense that there is some underlying change that must be discovered.***

Perceptual decisions are those in which the aim of the decision-maker is to categorize ambiguous (or noisy) sensory information

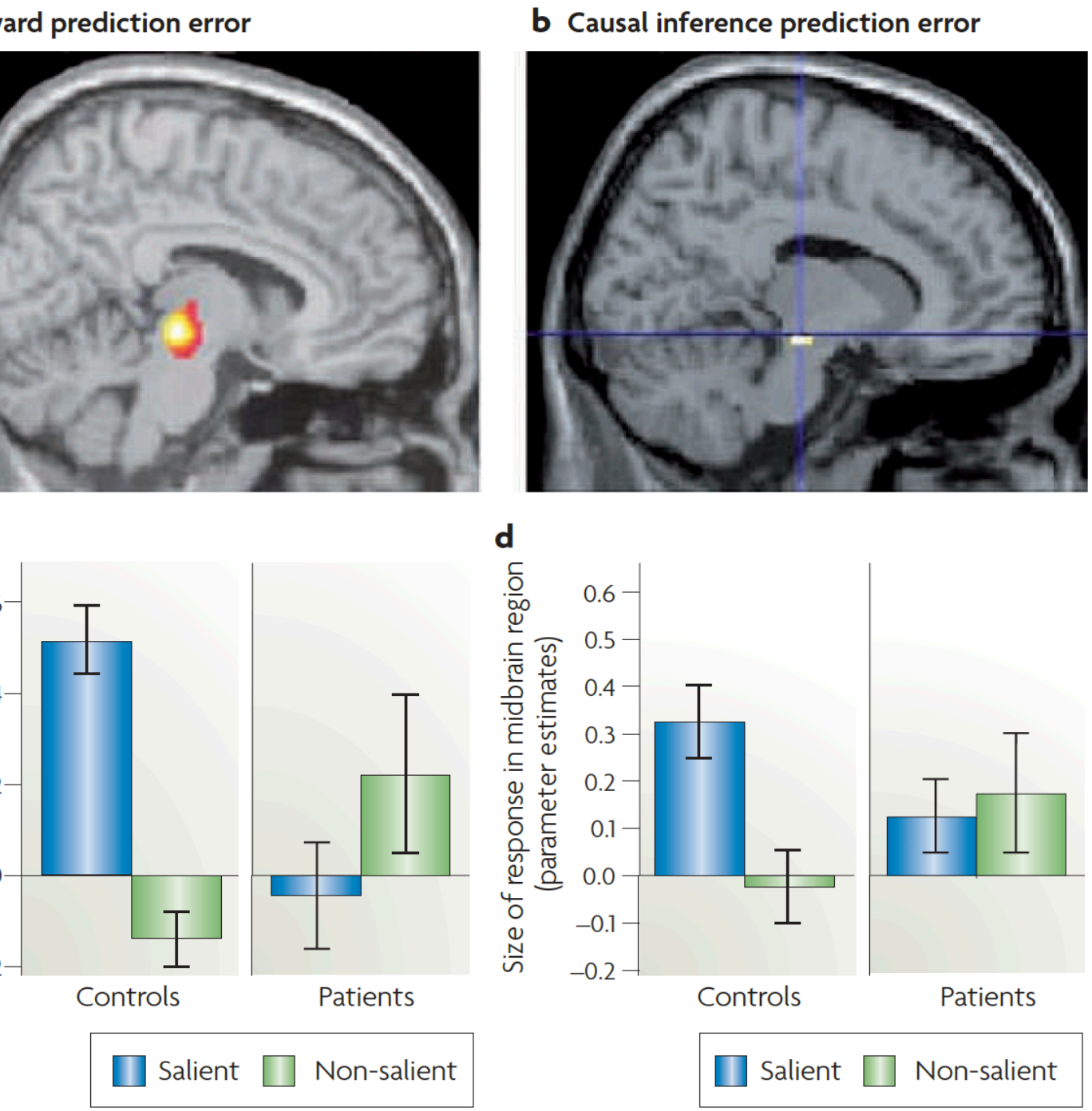
Murray, G. K. et al. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Mol. Psychiatry 13, 267–276 (2007).

Evidence of abnormal neural correlates of prediction errors in schizophrenia



Corlett, P. R. et al. Disrupted prediction error signal in psychosis: evidence for an associative account of delusions. Brain 130, 2387–2400 (2007).

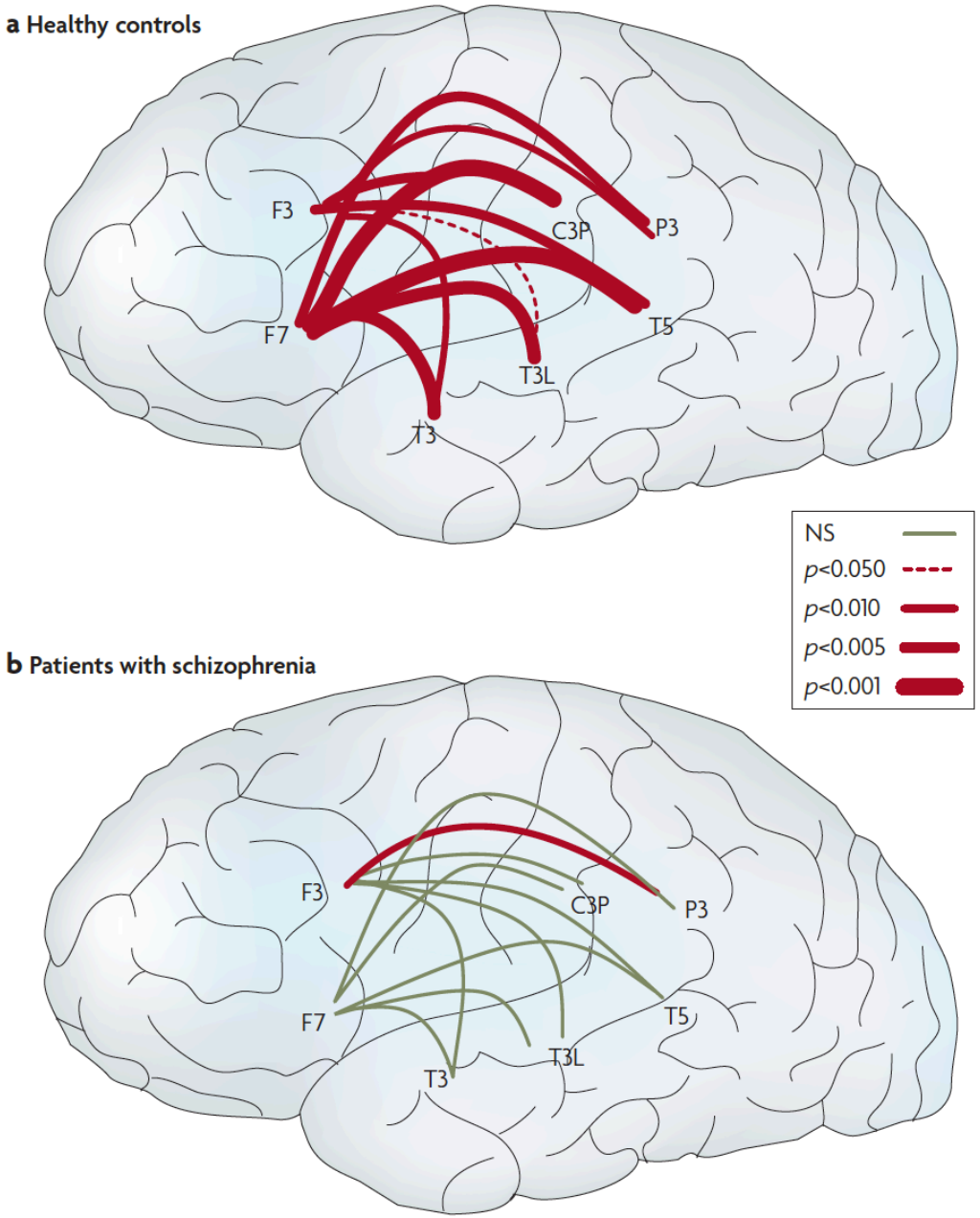
Delusions are maladaptive beliefs about the world. Based upon experimental evidence that prediction error mismatch between expectancy and outcome drives belief formation, this study examined the possibility that delusions form because of disrupted prediction-error processing. We used fMRI to determine prediction-error related brain responses in 12 healthy subjects and 12 individuals (7 males) with delusional beliefs. **Frontal cortex responses in the patient group were suggestive of disrupted prediction-error processing.** Furthermore, across subjects, the extent of disruption was significantly related to an individual's propensity to delusion formation.
Our results support a neurobiological theory of delusion formation that implicates aberrant prediction-error signaling, disrupted attentional allocation and associative learning in the formation of delusional beliefs.



Abnormal response to saliency in midbrain regions of patients with schizophrenia: functional MRI blood-oxygen-level-dependent (BOLD) activity in the midbrain.

All data were obtained from a group of 14 people with early psychosis and matched controls. (a, b) Functional MRI activity in the midbrain during tasks involving saliency. In one of the tasks **(a)**, stimuli acted as probabilistic indicators of either **financially rewarding ('salient')** or neutral ('non-salient') outcomes. In the other **(b)**, stimuli were part of a **causal inferential task** and constituted either a violation ('salient') or a fulfilment ('non-salient') of previously learned causal associations. (c, d)The size of the effect on the midbrain activation across groups and conditions for each of the experiments described above (c corresponds to a and d corresponds to b). ***In both experiments, control subjects' activation was greater for salient than for non-salient events, but this effect was attenuated and/or partially reversed in patients***

Fletcher and Frith (2009), Nature Reviews Neuroscience 10: 48-57



Abnormal connectivity associated with hallucinations.
Lateral views of the left hemisphere of the brain. The red lines connect areas that exhibited greater frontotemporal electroencephalogram coherence during talking than during listening for normal controls and patients with schizophrenia. The thickness of the line indicates the probability level for the t-tests that compared the findings. The thicker the line, the larger the difference between the two coherences. In **the controls, coherence during talking was was greater than during listening** for all 20 of the electrode pairs. In the **patients, coherence during talking was greater for only two of the pairs (one in each hemisphere).**
NS, not significant

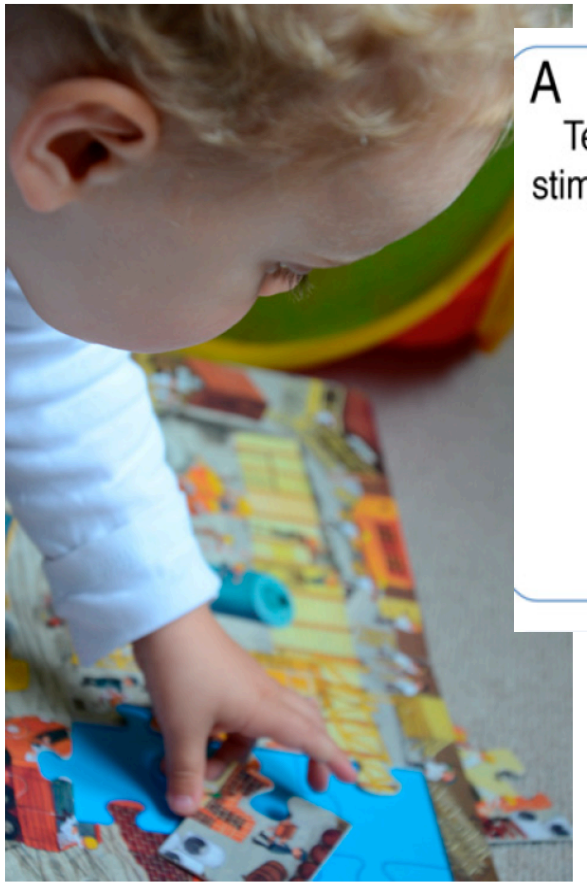
Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O. & Roth, W. T. Reduced communication between frontal and temporal lobes during talking in schizophrenia. Biol. Psychiatry 51, 485–492 (2002).

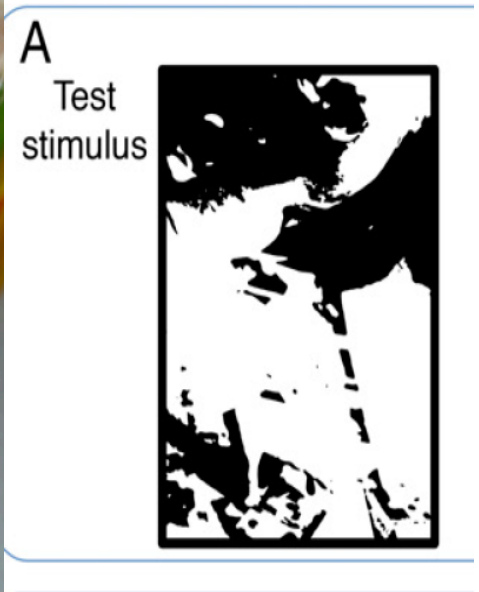
Fletcher and Frith (2009), Nature Reviews Neuroscience 10: 48-57

Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals

Christoph Teufel[a,b,1], Naresh Subramaniam[b], Veronika Dobler[c,d], Jesus Perez[c,d], Johanna Finnemann[b,e], Puja R. Mehta[b], Ian M. Goodyer[c,d], and Paul C. Fletcher[b,d]

PNAS vol. 112 | no. 43 | 13401–13406





Identified **unmedicated individuals who experience early psychotic symptoms but fall below the threshold for a categorical diagnosis**.
We observed that, in early psychosis, there was a shift in information processing favoring prior knowledge over incoming sensory evidence.
In the complementary study, we capitalized on subtle variations in perception and belief in the general population that exhibit graded similarity with psychotic experiences (schizotypy). We observed that the ***degree of psychosis proneness in healthy individuals, and, specifically, the presence of subtle perceptual alterations, is also associated with stronger reliance on prior knowledge***.

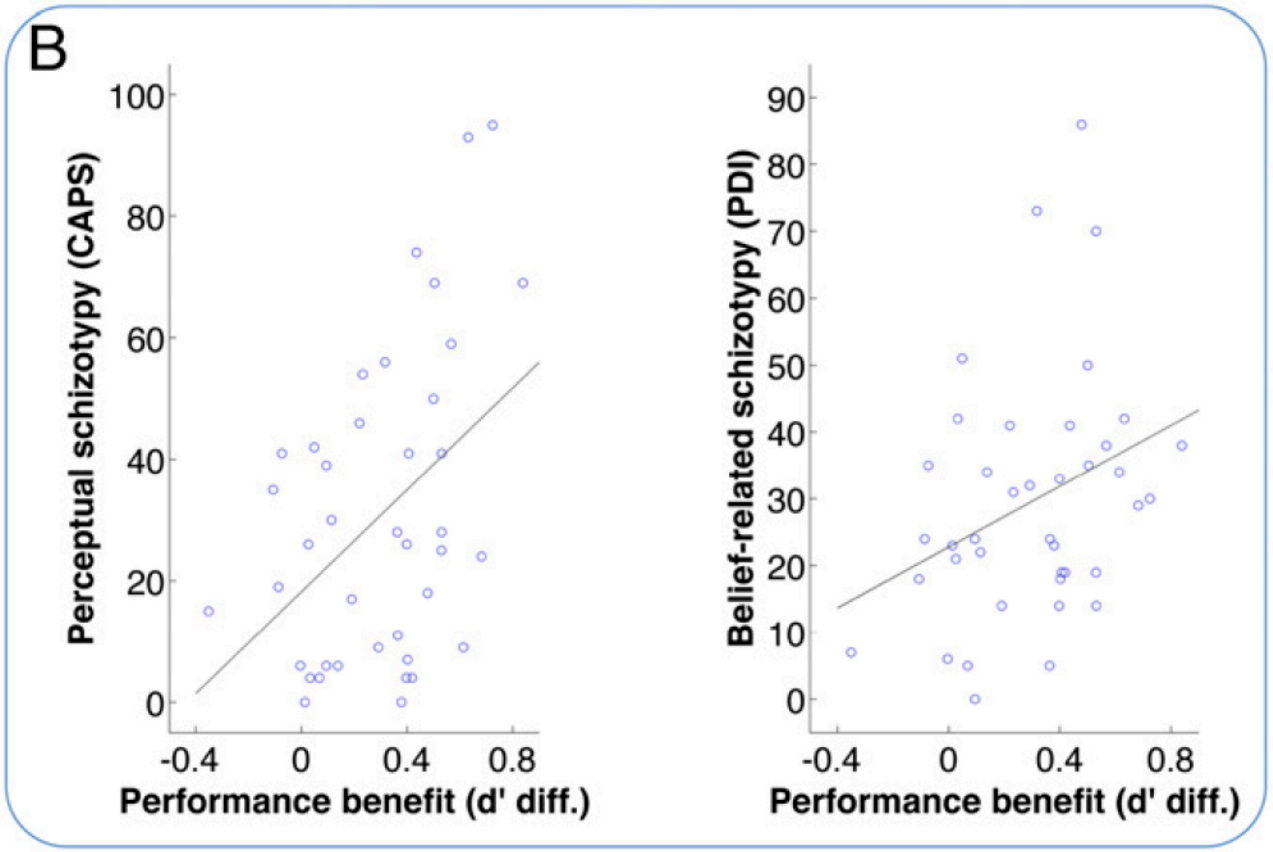
Perceptual decisions are those in which the aim of the decision-maker is to categorize ambiguous (or noisy) sensory information

Identified unmedicated individuals who experience early psychotic symptoms but fall below the threshold for a categorical diagnosis.

We observed that, **in early psychosis, there was a shift in information processing favoring prior knowledge over incoming sensory evidence**.

In the complementary study, we capitalized on subtle variations in perception and belief in the general population that exhibit graded similarity with psychotic experiences (schizotypy). ***We observed that the degree of psychosis proneness in healthy individuals, and, specifically, the presence of subtle perceptual alterations, is also associated with stronger reliance on prior knowledge***

Identified unmedicated individuals who experience early psychotic symptoms but fall below the threshold for a categorical diagnosis.
We observed that, **in early psychosis, there was a shift in information processing favoring prior knowledge over incoming sensory evidence**.
In the complementary study, we capitalized on subtle variations in perception and belief in the general population that exhibit graded similarity with psychotic experiences (schizotypy). ***We observed that the degree of psychosis proneness in healthy individuals, and, specifically, the presence of subtle perceptual alterations, is also associated with stronger reliance on prior knowledge***



Relation of performance with schizotypy. (Left) Performance benefit due to prior knowledge against the global CAPS score, which captures perceptual alterations that show graded similarity with perceptual atypicalities experienced during psychosis. Performance benefit was calculated as the difference in d′ between After and Before Blocks. (Right) Relation between performance benefit and global PDI scores of the participants. PDI captures belief-related alterations similar to those seen during psychosis.

Our studies were designed to characterize, in complementary ways, **the balance between visual bottom-up and top-down processing** in clinical individuals with early psychosis and healthy people prone to developing psychotic symptoms. A relative advantage in using prior knowledge to discriminate between ambiguous images was observed in both situations. ***This finding is especially striking in the clinical group in study 1 given that performance in this group (as in psychiatrically ill individuals more generally) is typically impaired. Such a result is rare and revealing in that it highlights a specific information-processing atypicality rather than a general performance deficit***

Visual function in early psychosis and in healthy people who are prone to such experiences is characterized by a basic information processing **shift that favors existing knowledge over incoming sensory evidence.** Although, *in the current experimental task, this shift conferred a performance benefit,* **under most natural viewing situations, it may provoke anomalous perceptual experiences. Specifically, it might impose prior expectations on inputs to the extent that, ultimately, formed percepts are generated that have no direct sensory cause: hallucinations**

Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making

Lorenz Deserno[a,b,c,1], Quentin J. M. Huys[d,e], Rebecca Boehme[c], Ralph Buchert[f], Hans-Jochen Heinze[a,b,g], Anthony A. Grace[h,i,j], Raymond J. Dolan[k,l], Andreas Heinz[c,m], and Florian Schlagenhauf[a,c]

PNAS | **February 3, 2015** | vol. 112 | no. 5 | **1595–1600**

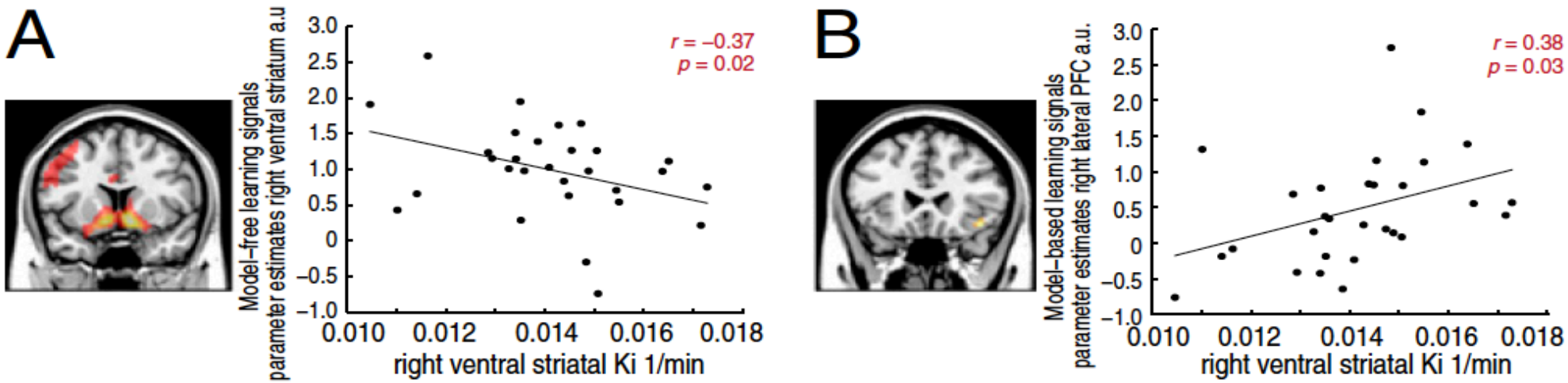


Fig. 3. Presynaptic dopamine and neural learning signatures. Correlation between right ventral striatal presynaptic dopamine K_i and (*A*) model-free learning signals in right ventral striatum ($r = -0.37$; $P = 0.02$) and (*B*) model-based signatures in right lateral prefrontal cortex ($r = 0.38$; $P = 0.03$).

Our studies were designed to characterize, in complementary ways, **the balance between visual bottom-up and top-down processing** in clinical individuals with early psychosis and healthy people prone to developing psychotic symptoms. A relative advantage in using prior knowledge to discriminate between ambiguous images was observed in both situations. This finding is especially striking in the clinical group in study 1 given that performance in this group (as in psychiatrically ill individuals more generally) is typically impaired. Such a result is rare and revealing in that it highlights a specific information-processing atypicality rather than a general performance deficit

Visual function in early psychosis and in healthy people who are prone to such experiences is characterized by a basic information processing **shift that favors existing knowledge over incoming sensory evidence.** Although, *in the current experimental task, this shift conferred a performance benefit, under most natural viewing situations, it may provoke anomalous perceptual experiences. Specifically, it might impose prior expectations on inputs to the extent that, ultimately, formed percepts are generated that have no direct sensory cause: hallucinations*