

OPINION

Adaptive learning under expected and unexpected uncertainty

Alireza Soltani¹ and Alicia Izquierdo²

Abstract | The outcome of a decision is often uncertain, and outcomes can vary over repeated decisions. Whether decision outcomes should substantially affect behaviour and learning depends on whether they are representative of a typically experienced range of outcomes or signal a change in the reward environment. Successful learning and decision-making therefore require the ability to estimate expected uncertainty (related to the variability of outcomes) and unexpected uncertainty (related to the variability of the environment). Understanding the bases and effects of these two types of uncertainty and the interactions between them — at the computational and the neural level — is crucial for understanding adaptive learning. Here, we examine computational models and experimental findings to distil computational principles and neural mechanisms for adaptive learning under uncertainty.

Imagine while in traffic that we decide on the route to our destination on the basis of commuting times experienced over many days, months or even years. Experiencing random small delays should not be concerning or prompt us to change our route. Yet unexpectedly slow traffic can signal important events (such as accidents or road closures), and this information should be used to update our route. Importantly, what we might consider to be unexpectedly long delays are very different for the dynamic metropolitan Los Angeles area versus the unchanging small town of Hanover. Nonetheless, successful learning requires mechanisms to discriminate inconsequential expected variability (which may lead to ‘expected uncertainty’) from signals of environmental volatility (which often leads to ‘unexpected uncertainty’), which should instead lead to a substantial update of value and/or changes in behaviour (such as a change in route). BOX 1 provides more complete, mathematical definitions of expected and unexpected uncertainty related to reward outcomes, but the definition of expected uncertainty can be summarized as the uncertainty (in terms of variance or standard deviation) attributable to the probabilistic nature of

reward outcomes, even when the probability of each outcome is fixed over time. By contrast, unexpected uncertainty may be defined as the uncertainty that is attributable to subjectively perceived volatility in the environment, which may differ from actual volatility or the rate of change in the environment over time.

Computing and updating the expected value of an outcome that follows selection of a stimulus or performance of an action (often referred to as stimulus values and action values) require the integration of signals across multiple brain areas and systems. These include areas involved in making decisions, executing actions and responding to primary rewards and their motivational and hedonic importance¹. In addition, updating stored information about reward could involve neuromodulatory systems that support neural plasticity. Some of the signals that influence the extent to which stimulus or action values are updated should be sensitive to uncertainty in the environment. However, learning in dynamic environments is bounded by a trade-off between adaptability (that is, the ability to respond quickly to changes in the environment) and precision (which requires slower update after each feedback)². One

way to improve this trade-off^{3,4} and allow more adaptability and precision at the same time may be to increase the rate of learning after unexpected events and decrease it when the world is stable.

Understanding learning under uncertainty requires an appreciation of how expected and unexpected uncertainty are computed, interact and, in turn, influence learning. In turn, understanding these types of uncertainty necessitates an examination of their relationship with stimulus or action values and whether and how they may be generalized across stimuli and actions to control an overall rate of learning. In this Opinion article, we compare existing computational models of different types of uncertainty and highlight their strengths and limitations. We examine experimental findings in the context of the predictions and basic principles of the computational models and discuss studies that implicate a distributed network of brain regions in computing expected and unexpected uncertainty. We distil testable hypotheses about how unexpected and expected uncertainty may interact and affect learning, and we offer ideas for future directions to pinpoint the neural substrates and mechanisms of adaptive learning. To make the topic more tractable, we do not discuss perceptual uncertainty or model-based uncertainty (uncertainty caused by changes in the model of the environment), as these add further layers of complexity and have been reviewed elsewhere^{5–9}.

Expected versus unexpected

Expected uncertainty. Specific definitions of expected uncertainty in the laboratory typically depend on both the nature of the learning or decision-making task and the model used by the organism to perform the task. In the case of estimating reward expected from a stimulus or action associated with different amounts of reward (outcome m_i) with different probabilities (p_i), expected uncertainty is equal to the variance (or standard deviation) of reward outcome in terms of magnitude or delay (BOX 1). The notion of uncertainty as the variance of probabilistic reward outcomes can also be applied to the perceived ‘risk’ of not getting reward when probabilities are known^{10,11}. However, we limit expected

Box 1 | Key terms and definitions

Below, we define certain key terms used throughout the main article.

Reward environment

We define 'reward environment' as a collection of stimuli and actions in which selection of a stimulus or execution of an action based on the presented stimuli brings about reward with certain magnitudes and probabilities. Reward obtained after selection of a stimulus allows a value to be assigned to the stimulus (stimulus value). Reward obtained after execution of an action can promote the formation of a stimulus–action association or the assignment of value to the action (action value). Reward attributes (such as reward magnitude and probability) can remain fixed or change over time, resulting in a stable or volatile environment, respectively.

Learning rate

'Learning rate' can be defined as the rate at which stimulus or action values are updated after each reward feedback. In error-driven models, the learning rate is a parameter between 0 and 1 that is multiplied by the reward prediction error (RPE) to determine the size of update or, equivalently, to determine how the observed reward should be weighted relative to previous stimulus or action values to update these values. In more mechanistic models, the learning rate can be seen as the rate of transition between different synaptic states^{33,106,107}. Dynamic environments require learning to be adjusted constantly, rendering the concept of 'learning rate' futile. Instead, we suggest referring to the modulation of the overall amount of update as the 'gain' of learning.

Expected uncertainty

'Expected uncertainty' is the uncertainty in reward outcome attributable to its probabilistic nature, even when the probability of different outcomes is fixed over time. For a stimulus or action that results in reward m_i with probability p_i , the expected uncertainty can be defined as the variance over n possible outcomes:

$$\text{Expected uncertainty} = \sum_{i=1}^n p_i \times (m_i - EV)^2,$$

where EV is the expected value.

$$EV = \sum_{i=1}^n p_i \times m_i$$

In the simple case of binary reward (reward m with probability p_r and zero otherwise), the variance in outcomes, or expected uncertainty, is equal to $p_r \times (1 - p_r) \times m^2$. The corresponding standard deviation can also be estimated by the average absolute deviation from the mean:

$$\sum_{i=1}^n p_i \times |m_i - EV|$$

Importantly, for binary outcome and $m = 1$, the mean absolute deviation can be computed by averaging the unsigned RPE over a large number of trials because the best estimate of EV is the expected reward resulting from stimulus or action selection. Therefore, the unsigned RPE may directly contribute to the computation of expected uncertainty for a given stimulus or action.

Volatility

Volatility refers to uncertainty due to 'actual' changes in reward magnitude and/or probability associated with stimuli or actions over time. In the context of the two-alternative probabilistic reversal learning task in which the probabilities for two possible actions add up to 1, it is proportional to

$$\frac{(2p_r - 1)}{L}$$

where L is the block length, capturing the overall rate of change per time. Volatility can be local (that is, related to one stimulus or action) or global (that is, generalized and shared between sets of stimuli or actions or all stimuli or actions in the environment). The computation of local volatility is reasonable because unexpected changes about one stimulus or action could be independent of those in other stimuli or actions. On the other hand, estimates of the overall level of volatility in the environment are beneficial in adjusting learning and decision-making on a global basis.

Unexpected uncertainty

Unexpected uncertainty is uncertainty due to subjective perceived changes in reward probabilities, magnitudes and/or delays that are associated with stimuli or actions over time. Unexpected uncertainty can be read only from a subject's responses (choice behaviour, estimation report and so on) and thus strongly depends on the assumptions of the computational model used to explain their behaviour. Unexpected uncertainty can be local or global.

uncertainty to cases in which probabilities must be estimated and do not change over time — that is, under stable conditions or in stable environments. It is deemed 'expected' because it is thought to reflect variability or stochasticity that is ubiquitous and unavoidable. In theory, encountering this kind of uncertainty should neither be surprising nor promote learning or behavioural adjustment over time; however, in practice, it is difficult for the subject to verify that probabilities do not change over time.

In most error-driven models of learning, the overall update of stimulus or action values depends on the product of the reward prediction error (RPE; the difference between the expected value and the outcome) and the so-called learning rate (BOX 1). Because RPE is determined independently of the environment, the continuous adjustment in the overall amount of learning required in dynamic environments is often described using time-dependent learning rates. However, such descriptions are somewhat futile conceptually, as they are analogous to defining the overall speed for a pendulum, which has time-dependent speed. In addition, it is unclear how such learning rates are determined at the neuronal level. Therefore, instead we refer to the 'gain' of learning in lieu of the learning rate to describe the modulation of the overall amount of update in stimulus or action values. Nevertheless, expected uncertainty has been suggested to scale the learning rates to reduce the influence of prediction error when reward outcomes are more variable^{12,13}. This strategy is useful only if the environment is stable enough that the variance can be estimated reliably, as a large variance could also reflect a real change in the environment, which should instead enhance learning. One way of computing an estimate of expected uncertainty is to average the unsigned RPE over time (that is, over trials), because the latter can approximate the standard deviation of reward outcomes. Thus, rather than directly scaling down the gain of learning, expected uncertainty could indirectly influence learning by providing a 'baseline' level of variability against which surprising events that should increase the gain of learning can be detected (see below).

Unexpected uncertainty. Unexpected uncertainty occurs owing to changes in reward probabilities, magnitudes and/or delays over time when taking expected uncertainty about the variability of

outcomes into account; however, its strict definitions differ^{14–16}. For example, it has been suggested that only the first violation of an expected outcome, such as with the first reversal of previously learned reward contingencies, constitutes unexpected uncertainty, whereas all subsequent reversals yield expected uncertainty because they could be expected^{14,17}. This posits that there is no unexpected uncertainty if changes are predictable; however, this principle does not add much practical benefit to the definition of unexpected uncertainty because it is not clear when the ‘unexpectedness’ of changes degrades. Learning can occur in a single trial, as in epiphany learning¹⁸, suggesting that detection or perception of drastic changes in the environment that lead to such learning can be very fast. Thus, some have argued that to capture ‘surprise’ correctly, commitment to a belief needs to be considered as well¹⁶. Unexpected uncertainty may therefore be primarily subjective rather than following actual changes in the environment.

Thus, to reconcile different definitions, we equate unexpected uncertainty with the ‘subjective’ perceived uncertainty due to changes in reward probabilities, magnitudes or delays over time (BOX 1). We assert that unexpected uncertainty can be read out only from a subject’s responses (for example, from their choice behaviour or estimation report). By contrast, we refer to volatility as uncertainty due to actual changes in reward probabilities, magnitudes or delays over time, independently of whether it is detected by the decision-maker or not. Given that unexpected uncertainty strongly depends on the assumptions of the computational model used to explain subjects’ behaviour (and given that subjects adopting such a model could require substantial time), whether volatility is computed and signalled by neural elements in the brain — and, if so, how it influences learning — must be tested.

Computational models

Here, we categorize computational models of learning under uncertainty into normative models, such as Bayesian or statistical models, that prescribe how learning should adjust to uncertainty in the environment; approximations to these normative models that aim to provide plausible update rules that could be implemented in the brain; and last, but not least, mechanistic models that aim to explain how necessary computations can be performed by neural elements.

Normative models. An ideal observer in Bayesian and statistical models of learning uses the Bayes rule as the update rule to estimate the probability of reward or other important quantities optimally^{19–21}. However, this requires making certain assumptions about the environment to determine what regularities to expect and to learn from in the environment^{5,6,15,17,19,20,22,23}. Therefore, the decision-maker or learner assumes a model of the environment and how it changes over time and accordingly adjusts the parameters of this model on the basis of reward feedback. Parameters of the Bayesian models could represent different properties of the environment such as the probability of reward, the width of distribution from which reward is drawn (expected uncertainty) and the probability that any of the underlying parameters may change over time (unexpected uncertainty)^{15,17,19,20,22}. Therefore, Bayesian models estimate not only stimulus or action values but also expected and unexpected uncertainty associated with those values and are therefore very useful for localizing corresponding neural correlates.

Many Bayesian models of learning under uncertainty assume a hierarchical structure for estimating relevant reward information (probability, magnitude and so on) and how it changes over time^{15,20,22}. This assumption is usually made for mathematical convenience and may not reflect the type of changes that occur in the natural reward environment. One hierarchical model²⁰ posited separate systems for estimating reward probability (r), volatility (v) and the rate of change in volatility (k). In this model, transitions between different values of r and v are affected by the parameter in the system above it (v and k , respectively). In this structure, the Bayes rule can be applied to compute posteriors (posterior probability distributions) or the belief about all three parameters given the data. This model successfully explained choice behaviour in a probabilistic reversal learning task and was used to identify neural correlates of unexpected uncertainty in humans; moreover, this model sparked the development of many models of learning under uncertainty. Despite its logical simplicity, the actual computations necessary for estimating posteriors are very complex; therefore, it is unclear precisely how these computations are performed in the brain. Most relevant to our discussion, it is difficult to use Bayesian models to make predictions about the exact relationship between expected and unexpected uncertainty as their update rules are closely interconnected²².

A common normative approach for tackling learning under uncertainty is the Kalman filter. The Kalman filter utilizes series of observations to produce estimates or predictions for both the state of the system and uncertainty associated with the estimated state, and it uses subsequent observations to update those estimates. As a result, the Kalman filter lends itself well to learning under uncertainty because it formalizes not only the predictive relationship between stimuli or actions and reward outcomes (predicted state, such as reward probability) but also the variance or uncertainty of the estimates (so-called predicted error covariance)^{5,6}. An important concept in this model is the optimal Kalman gain, which determines the amounts of update for both the predicted state and the predicted error covariance⁵. Similar to Bayesian models, Kalman filter models require an assumption about state transitions (more specifically, a state-transition model), and the most common form of state transitions follow a hierarchical structure^{5,6}. Both Bayesian and Kalman filter models have been instrumental in formalizing different solutions to tackle uncertainty.

Approximate-normative models. Although optimal and quite generalizable, the computations required in the normative models are rather complex and cannot easily be mapped to neural processes. Moreover, because normative models are mainly concerned with describing optimal learning, these models are sometimes limited in accounting for choice and learning behaviours^{2,15}. Different approaches have been used to overcome these issues. These include approaches incorporating additional components to the Bayesian models or approximations to those models. For example, Payzan LeNestour and Bossaerts¹⁵ proposed a ‘forgetting’ Bayesian algorithm, enabling the introduction of an explicit learning rate. Using this model, the authors were able to estimate the effects of different types of uncertainty on the learning rate and identified multiple brain regions (including the anterior and posterior cingulate cortex, intraparietal sulcus and locus coeruleus) that displayed blood-oxygen-level-dependent (BOLD) responses that were correlated with different types of uncertainty²⁴.

In another example, Dayan et al.⁵ noted that Kalman filter models assume that the predictive values of all stimuli are simply added to compute ‘net predictions’, although the reliability of those stimuli could vary owing to abrupt changes in the environment. To resolve this issue, they proposed a

‘competitive combination’ mechanism that uses the inverse of the standard deviation of the difference between the predicted and actual value of reward associated with each stimulus (as a measure of stimulus reliability) to optimally combine different predictions. To solve a similar problem related to differences in reliability, Courville et al.⁶ also suggested that the update of Bayesian model parameters or beliefs about them should inversely depend on uncertainty in the environment. However, they also proposed that surprising events or outcomes should signal changes and the need for faster or new learning.

There are other approximate Bayesian models that arguably provide better fit to behaviour and links to its neural substrates^{19,22,25}. For example, a model comprising a mixture of error-driven ‘delta’ rules that generated updates on the basis of the differences between estimated and actual outcomes explained human behaviour on a predictive-inference task more effectively than did the optimal Bayesian model²⁵. A related study proposed a delta-rule approximation of the ideal observer model¹⁹; in this model, the influence of newly experienced outcomes was adjusted according to ongoing estimates of (expected) uncertainty and the probability of a fundamental change in the environment (unexpected uncertainty). These approximate models tend to provide a good fit of behavioural data (particularly for continuous reward feedback, although not binary reward feedback) and have been used to identify neural correlates of belief updating²⁶. These models can be directly tested for their predictions about changes in the learning rates and how the effect of RPE on learning is modulated by environmental factors.

Classic models of learning based on RPE (for example, the Rescorla–Wagner model and various reinforcement learning (RL) models²⁷) assume fixed learning rates and thus do not have specific mechanisms for adjusting the learning rate according to uncertainty in the environment. By contrast, the Pearce–Hall (PH) model carries a built-in mechanism for adjustment of learning for each stimulus based on how surprising the last outcome was, such that surprising reinforcement (or non-reinforcement) results in increased associability and faster learning²⁸. This is why in many hybrid models that try to capture changes in learning over time, RPE is multiplied by a variable that measures surprise, which could also signal uncertainty²⁹. In the PH model, surprise is computed on the basis of the mean value of unsigned RPE whereas in

other models, this variable could resemble the variance of the RPE^{6,12}; nonetheless, all these models suggest that RPE can be used to control the gain of learning. For example, Preuschoff and Bossaerts¹² suggested that the standard deviation of the RPE, which they term the prediction risk, can be used to adjust the learning rates by scaling down the RPE. Another study proposes a model with a dynamic learning rate based on the slope of the change in the smoothed, unsigned RPE over trials³⁰. Interestingly, a recent model that combined these two features (scaling of learning with expected uncertainty and dynamic learning rates using RPE) was able to account for human learning better than classic RL models¹³. Although the notion of variable learning rates based on unsigned RPE has been adopted in many models to deal with uncertainty^{5,6,31}, RPE does not tease apart expected from unexpected errors per se. Therefore, additional computations are necessary for proper estimation of surprising outcomes, and the neural mechanisms of these additional computations are currently unknown.

Mechanistic models. Mechanistic models of learning under uncertainty aim to explain how necessary computations are performed by neural elements; thus, their components can be mapped onto brain circuits and substrates more easily than the components of the models outlined above.

In error-driven models, the RPE modulated by the learning rate is often used as the teaching signal, which is assumed to be mediated by dopamine^{32,33}. Therefore, adjustments of learning in these models translate to the adjustment of RPE or the learning rates or both. However, a wealth of evidence implicates dopamine in other processes that also influence choice behaviour, including incentive salience or desirability³⁴, effort³⁵ and novelty³⁶. Thus, whether the computations required for adjustments of learning in error-driven models can be mapped uniquely onto the modulation of the functionally multifaceted dopaminergic system is unclear.

We recently proposed a mechanistic model for adaptive learning under uncertainty in which synapses are endowed with metaplasticity — the ability to change synaptic states without measurable changes in synaptic efficacy. Through such changes, metaplasticity can alter future response to events that cause plasticity without producing any potentiation or depression³⁷. As such, synapses with reward-dependent metaplasticity can self-adjust in response to reward statistics in the environment without

any optimization or knowledge of the environment^{2,4}. In this model, the changes in the activity of neurons that encode stimulus or action values can be used by another system to compute volatility in the environment. The volatility signal can subsequently be used to increase the gain of learning when estimated volatility passes a threshold set by expected uncertainty. Therefore, the extended model predicts a direct two-way interaction between neurons encoding stimulus or action values and neurons computing volatility, modulated by input from a circuit that computes expected uncertainty. We posit that such interactions between value-encoding and uncertainty-monitoring systems can enhance the adaptability required in dynamic environments and that metaplasticity provides a crucial mechanism for this interaction to be beneficial. Metaplastic synapses are crucial because they allow proper estimation of volatility². In addition, according to this model, expected uncertainty may not directly modulate the gain of learning but instead may be involved in setting a baseline with which to compute unexpected uncertainty.

Another proposed model³ for learning under uncertainty consists of two networks. The first network exploits reward-based metaplasticity to estimate stimulus or action values. The second network updates plasticity in the first network according to the degree of ‘surprise’ it detects. The second network computes ‘surprise’ on multiple timescales by comparing the current differences in reward rates over pairs of timescales (referred to as unexpected uncertainty) with the means of these differences over time (referred to as expected uncertainty). Therefore, unlike our model², this model proposes that completely separate systems estimate stimulus or action values and compute uncertainty. In addition, this model predicts that the surprise detection system should have a unidirectional influence on the valuation system, whereas our model predicts two-way interactions between systems encoding stimulus or action values and volatility. These alternative predictions can be tested by pathway-specific inactivation of brain regions involved in computations of different types of uncertainty (see below) and measuring the effect on learning.

Neural substrates of uncertainty

Here, we examine recent findings from studies in the brain to identify evidence that may serve to validate or refute the predictions of the aforementioned computational models. With few exceptions³⁸, most experimental paradigms

probing the neural substrates of uncertainty rarely involve the clear distinctions between the different types of uncertainty studied here (BOXES 2,3). Therefore, in the discussion below, we focus on recent findings that provide insights on correlates, circuit mechanisms and connections that may compute, encode or distinguish between the uncertainty-related variables described above. We also consider studies that aim to reveal causal roles using different interference methods. Overall, these findings point to a distributed network that includes regions of the prefrontal cortex (PFC), striatum, hippocampus, basolateral amygdala (BLA) and mediodorsal thalamus (MD) (FIG. 1). Owing to space constraints, we focus on reward-based rather than aversion-based learning³⁹ and emphasize corticolimbic contributions to this learning rather than those of supporting neuromodulatory systems. Several studies and reviews have already outlined the important contributions of dopamine^{40,41}, serotonin^{1,42} and acetylcholine and noradrenaline¹⁴ to learning and decision-making under uncertainty.

Prefrontal cortex. Neural correlates of uncertainty have been found in various regions of the PFC, including the anterior cingulate cortex (ACC), of several species⁴³. ACC neurons were reported to represent unsigned RPE, which was correlated with behavioural adjustment in a gambling task⁴⁴. In a more recent study⁴⁵, ACC neurons were shown to signal both expected value and expected uncertainty in a valence-specific manner. Moreover, of the ACC neurons that signalled uncertainty, fewer signalled RPEs than variability in reward outcomes (expected uncertainty). Thus, ACC-mediated encoding of RPE, previously observed by several groups^{46–48}, could contribute to uncertainty computations in this area. Although encoding of RPE itself does not qualify a brain area for uncertainty computations, signals of the unsigned RPE can provide strong evidence for approximation of expected uncertainty.

Unlike the ACC, there is debate on whether the orbitofrontal cortex (OFC) signals RPEs^{49–51} and whether and how these signals may contribute to different

forms of uncertainty. Electrophysiological recording studies in the rat OFC provided convincing evidence that activity in this region correlates with both stimulus value and expected uncertainty^{52,53} similar to ACC. This ability to encode these variables depends on the stability of the environment⁵², suggesting a contribution of the OFC to expected uncertainty and perhaps even volatility⁵⁴. Most of the functions described above in rats⁵⁵ have been realized in the monkey brain as well: representations of expected outcomes can also be decoded from monkey OFC during value-based choice^{56,57}, and OFC neurons signal both stimulus value and expected uncertainty⁵⁸. In monkeys, the activity of OFC neurons rapidly updates in response to changes in cued reward magnitude, and activity in this region, as in the rat OFC, is modulated by reward history⁵⁹. For example, signals relevant to task performance can be decoded better in both the OFC and ACC in a volatile environment than in a stable environment⁶⁰. Together, the evidence in rat and monkey OFC points to conserved functions in learning under uncertainty.

Which of these PFC regions are causally involved in uncertainty computations? Anatomical and functional data support the idea that the ACC in various species integrates reward, cognitive and action plan signals^{61–65}. Similarly, lesions and transient or reversible pharmacological inactivation of the OFC in rats and monkeys also result in deficits in learning and performance under conditions of risk and uncertainty^{55,66–72}. An important follow-up question is how these regions contribute to uncertainty computations. The OFC (and perhaps also the ACC) has access to volatility signals such that they can construct stable representations of stimulus or action values under expected uncertainty^{52,68}. However, computations of volatility or unexpected uncertainty themselves probably occur outside the PFC in regions such as the BLA that are interconnected with both cortical and subcortical regions (FIG. 1).

On the basis of the available evidence, we speculate that the OFC and ACC may have dissociable and complementary roles in learning under uncertainty and in mitigating the adaptability–precision trade-off. Specifically, the OFC may support slow updates of stimulus values and estimation of expected uncertainty over multiple trials to provide a baseline for computing unexpected uncertainty. Plastic synapses in the ACC may instead have a spectrum of learning rates or rates of transition between potentiated and depressed states^{2,73}.

Box 2 | Behavioural paradigms used to study expected and unexpected uncertainty

Learning and decision-making paradigms with probabilistic outcomes that manipulate the variance in reward outcome can be used to study expected uncertainty and its influence on learning^{66,108}. A popular approach for studying the effects of expected uncertainty on learning and choice is to require animals to select between a probabilistic large-magnitude reward (the probability of which must be learned) and a certain small-magnitude reward^{67,109–111}. A related concept is risk, which corresponds to expected uncertainty owing to the probabilistic nature of the outcome in cases in which reward probabilities are provided and not to be learned.

Another way in which some groups have attempted to study expected uncertainty is to introduce variability in reward magnitudes or delays to reward^{13,19,58,68,112}. In such paradigms, subjects are typically required to select between stimuli associated with different delay variance or to estimate reward magnitudes for stimuli associated with rewards with different probability distributions (but with the same mean reward and reward rate over the session). Evidence suggests that non-human primates^{19,58} and rodents^{68,112} can infer the standard deviation of a reward distribution (expected uncertainty).

Most experimental paradigms used to study unexpected uncertainty involve learning of frequently changing associations between stimuli or actions and reward outcomes^{17,21,60,113–115}. A commonly used paradigm is probabilistic reversal learning (PRL), in which the subject selects between two alternative options (for example, visual stimuli) that each result in probabilistic delivery of reward. The probabilities of reward on the two options can switch after a certain number of trials (L ; or block length) that could be fixed^{20,114} or drawn from a distribution²¹. Alternatively, a switch in reward probabilities could be determined on the basis of when performance reaches a certain level, but such design is not ideal because it makes volatility contingent on choice. In the case of fixed block length L , a combination of reward probability on the two options and L collectively defines the reward environment, with a specific value of volatility (BOX 1). PRL is a challenging task for two main reasons: the probabilistic nature of reward assignment (expected uncertainty) and the frequent switches in reward probabilities between blocks of trials (reversals), which result in volatility and thus unexpected uncertainty. In more general ‘bandit tasks’, the subject selects between two or more options or actions for which reward probabilities can independently change over time^{116–119}. Although bandit tasks can introduce different forms of uncertainty, they are not the most suitable for dissociating them or studying their influence on learning for several reasons. First, reward probabilities change on the basis of independent, random processes for different options, making it difficult to define volatility or unexpected uncertainty. Second, outcomes for the unchosen options or actions are unobservable; thus, estimating reward probabilities for these options is non-trivial. Last, owing to such unobservable outcomes, bandit tasks involve the use of complex strategies (such as exploration–exploitation trade-off) for selecting recently unchosen options or actions.

PERSPECTIVES

This would not only enable the ACC to estimate expected uncertainty but also allow computation of unexpected uncertainty elsewhere in the brain. Thus, the OFC and the ACC could provide parallel signals necessary for computations of unexpected uncertainty elsewhere, where these signals can be compared and used for faster updates if unexpected events are detected.

Does the functional connectivity support this possibility? Anatomical studies in rodents and monkeys point to a topographic map of connectivity from various subcortical and cortical structures to the PFC⁷⁴. Specifically, along the lateral-to-medial gradient in the rat OFC, innervation by affective or motivational systems increases and innervation by sensory integration areas

decreases (reviewed in REF.⁵⁵). In the ACC, there is a similar pattern of connectivity but along the ventral–dorsal plane, with ventral areas connected to the amygdala and dorsal areas better connected with sensorimotor and association areas⁷⁵. This connectivity results in the transfer of largely redundant information to the OFC and ACC that could be used to compute different quantities, such as stimulus or action values, the unsigned RPE and others, that are required for uncertainty computations. The role of corticocortical connectivity and interactions in forming representations of value and uncertainty should not be neglected. The medial and ventral OFC send dense projections to the dorsal and ventral ACC⁷⁶, suggesting that the ACC receives both

direct and ‘OFC-filtered’ information about rewards and may therefore represent reward information differently to the OFC⁷⁷. Moving forward, it will be crucial to discern the conditions and timings with which the ACC and OFC may be differentially engaged in learning about reward and to compare their contributions directly on the same tasks.

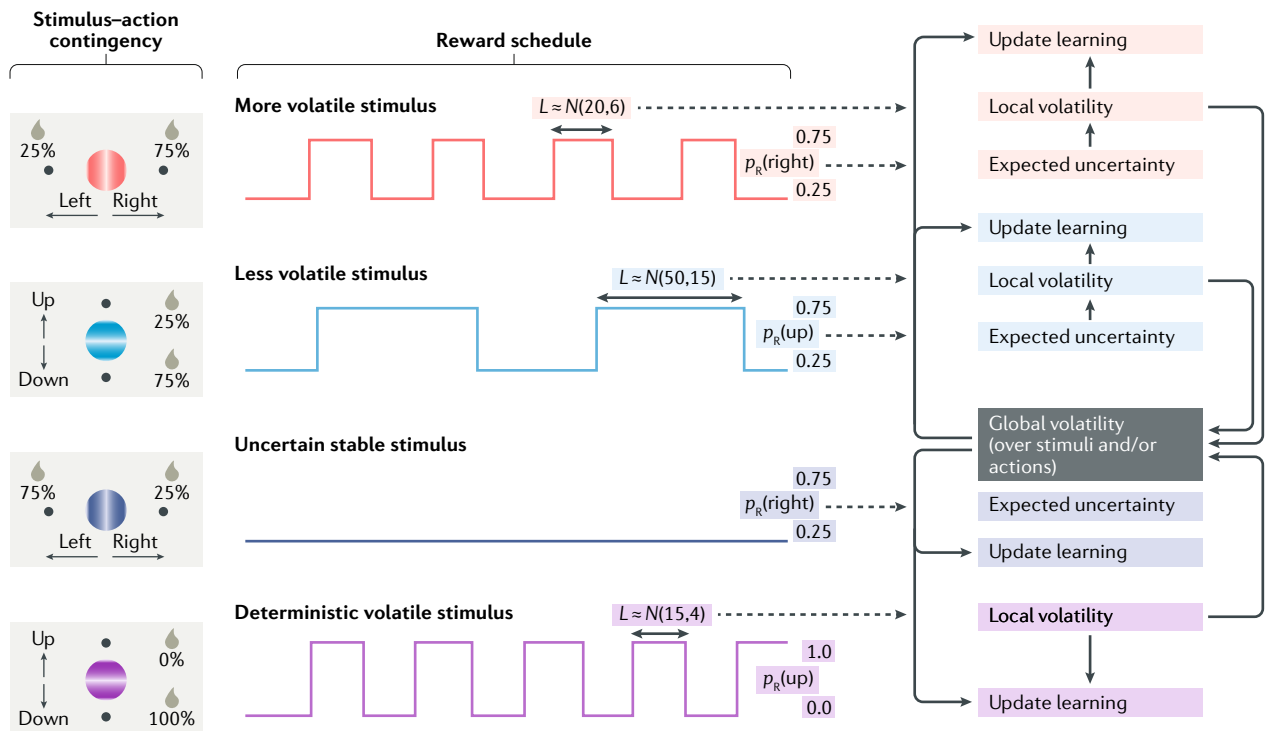
Striatum. Correlates of expected uncertainty (as defined here) and stimulus–outcome associations have been found in the dorsal striatum in monkeys⁷⁸. Evidence also suggests that the striatum may also be causally involved in learning under expected uncertainty. Lesion studies support a role for the striatum in learning during probabilistic, rather than deterministic,

Box 3 | Example experimental paradigm to dissociate expected and unexpected uncertainty

Components of probabilistic reversal learning tasks can be used for a new experimental paradigm to dissociate expected and unexpected uncertainty and to study their interaction. In this task, the subject concurrently learns stimulus–action associations for multiple visual stimuli via reward feedback, as in a natural environment. Stimuli could have similar features and could be associated with similar or different sets of actions (see the figure). The reward outcomes for actions associated with each stimulus can be probabilistic or deterministic and can reverse on a specific timescale (or block length (L), drawn from a normal distribution with a specific mean and variance, for example, $N(20,6)$).

A stimulus with a fixed, probabilistic stimulus–action association (no reversal) carries expected uncertainty but not unexpected uncertainty. A stimulus with deterministic outcomes and reversal bears only unexpected uncertainty. Other stimuli with probabilistic reward outcomes and reversal give rise to both types of uncertainty. The neural responses to these three types of stimulus can be used to dissociate the neural correlates of different types of uncertainty.

Moreover, the subject could construct models of the environment to predict reversals, which gives rise to the distinction between ‘objective’ volatility and ‘subjective’ unexpected uncertainty (BOX 1). In addition to stimulus-specific (or ‘local’) volatility or unexpected uncertainty, global volatility or global unexpected uncertainty could be computed across stimuli and/or actions to determine the overall gain of learning in the environment. To compute global volatility, volatility associated with each stimulus or set of actions can be summated with a weight proportional to the inverse of the expected uncertainty. These alternatives can be dissociated using different levels of similarity between stimuli and between sets of actions. Finally, this paradigm allows studying the interaction between expected and unexpected uncertainty in terms of how expected uncertainty is used to compute volatility (or unexpected uncertainty) about each stimulus or sets of actions; whether and how expected uncertainty is used to combine volatility across stimuli and/or sets of actions; and how expected and unexpected uncertainty influence learning.



reward schedules in both rodents and non-human primates^{79,80}. The striatum receives inputs about expected and unexpected uncertainty from frontocortical regions, the hippocampus and the BLA (FIG. 1) and is necessary for tasks that require encoding of reward rate and delay to reward^{79,81}, thereby implicating this region in flexible learning (and responding to) changes in individual outcomes.

Hippocampus. Expected uncertainty signals associated with probabilistic outcomes have been reported in the septum⁸², which in turn may aid learning via its innervation of GABAergic interneurons in the hippocampus⁸³. Conversely, unexpected uncertainty correlates have also been found in the hippocampus, mostly reported in primates. For example, in humans, correlates of change detection and ‘mismatch’ computations were observed in the hippocampus⁸⁴. In addition, the amplitudes of negative event-related potentials in the hippocampus covary with unexpected uncertainty in outcome, irrespective of valence⁸⁵. Moreover, the hippocampal BOLD signal correlates negatively with unexpected uncertainty at the time of outcome²⁴. The hippocampus and OFC have both been suggested to support cognitive maps that could provide predictions about choice outcomes^{86,87}. Given this critical overlap in function between these two regions, the hippocampus may also conceivably signal expected uncertainty in addition to unexpected uncertainty. To our knowledge, there has been no direct test of a causal role for the hippocampus in learning under the different forms of uncertainty that we consider here, in rodents or primates.

Basolateral amygdala. A large body of data in rodents and non-human primates points to the BLA in detecting surprising changes, leading to quick updating^{88,89} that supports flexible learning^{79,90}. Indeed, BLA activity responds to changes in the internal (motivational) state of the animal, which is typically probed using reinforcer devaluation paradigms⁹¹. However, the BLA also encodes changes in the (external) environment as positive or negative RPEs when expectations are repeatedly violated^{91,92}. The ability of the BLA to encode changes in the environment has been modelled in terms of attentional salience and ‘associability’ signals (as in the PH model)²⁹, both of which could contribute to computations of unexpected uncertainty. For the BLA to facilitate rapid updating, this region must also receive information about expected uncertainty (FIG. 1).

More recent evidence indicates that the BLA supports learning of actual changes (that is, volatility) in the expected value of rewards as signalled externally⁶⁸, not necessarily by shifts in (internal) motivational state. The BLA may directly influence value learning under uncertainty via its anatomical projections to or from the ACC^{93,94} and OFC^{95,96} and/or via the dopaminergic circuitry⁹⁷. These projections could allow the BLA to compute unexpected uncertainty by comparing changes in stimulus or action values to baselines set by the expected uncertainty.

Mediodorsal thalamus. The MD is an important node for value processing^{98,99} but has also been explored for its involvement in rapid learning in changing environments^{100–102}. For example, monkeys with MD lesions exhibit an increased tendency to switch response, even after a win trial¹⁰¹, suggesting that MD is required for maintaining a representation of recent reward modulated by choice, which could facilitate learning when there are multiple stimuli in the environment. In support of this, one study¹⁰³ suggested that the architecture of thalamocortical and corticothalamic pathways may support the maintenance and rapid update of cortical representations, making MD a candidate region for volatility computations.

A network for computing uncertainty.

The experimental evidence for neural substrates for computations of uncertainty raises several important points. First, although certain regions of the brain seem to show some tendency (although are probably not ‘specialized’) for encoding and computing different variables related to uncertainty in the reward environment, computations and perhaps representations of uncertainty are distributed. Second, because stimulus and action values, the unsigned RPE and expected uncertainty are closely linked both anatomically and behaviourally, computations of uncertainty may not require separate estimation of stimulus and action values. In other words, the same stimulus and action value signals may be used to compute both types of uncertainty. Last, although there are connections from cortical areas involved in valuation and uncertainty computations to subcortical areas involved in signalling reward and prediction error, whether dopaminergic systems receive and can integrate different types of uncertainty information to modulate the learning rates, as has been widely proposed, remains unclear.

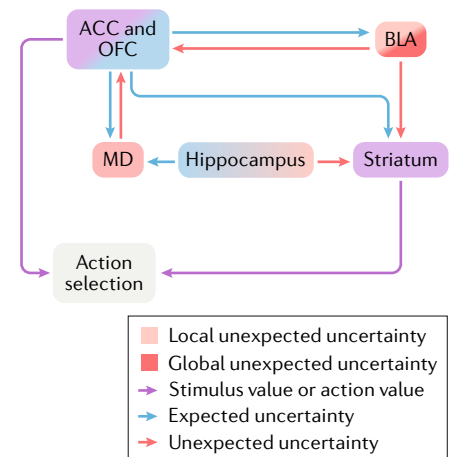


Fig. 1 | Major nodes of expected and unexpected uncertainty computations. On the basis of existing data, these are a few cortical and subcortical areas that could be involved in the computations (and representations) of expected and unexpected uncertainty as well as stimulus or action values. We do not include all anatomical connections for simplicity. The uncertainty network includes the anterior cingulate cortex (ACC), basolateral amygdala (BLA), hippocampus, mediodorsal thalamus (MD) and orbitofrontal cortex (OFC). Most of these areas are highly reciprocally interconnected, which could explain the overlap in the information or variables that each of these areas represent and compute and demonstrates that learning under uncertainty involves inherent interactions between expected and unexpected uncertainty signals.

Future perspectives

To allow effective learning, the brain must achieve a balance of ‘scaling down’ learning when expected uncertainty is high versus ‘scaling up’ learning when unexpected uncertainty is high. Above, our discussion of computational models and experimental data suggests that understanding the interactions between expected and unexpected uncertainty is crucial for understanding learning and choice under uncertainty.

Nevertheless, several issues remain to be addressed (see BOX 4 for remaining questions and how they can be addressed). Although neural correlates of expected uncertainty signals have been found in the brains of several different species, we still do not know how these signals contribute to the computations of unexpected uncertainty or volatility and subsequent learning. We also lack an understanding of how unexpected uncertainty is encoded in the brain (for example, whether it is encoded at the single-cell level or at the population level) or how it is computed, although metaplasticity may provide a promising mechanistic framework for its computation^{2,37}.

Box 4 | **Outstanding questions**

Several questions concerning the computation of expected and unexpected uncertainty remain unanswered (see below). Some of these questions can be tested currently by pathway-specific manipulations in rodents, multi-area recoding in non-human primates and behavioural manipulations using novel paradigms in humans and other species (BOX 3). In vivo imaging in target prefrontal cortex regions while specific pathways and cell populations are activated or silenced during learning would be especially revealing. In addition, answering these questions requires understanding interactions between neural elements across multiple levels (including at the synaptic and circuit levels) and will therefore require detailed computational modelling.

How do uncertainty signals interact across brain areas?

Exactly how uncertainty computations and corresponding signals in the anterior cingulate cortex (ACC) and orbitofrontal cortex (OFC) interact with those in the basolateral amygdala (BLA) and the mediodorsal thalamus (MD) to enable adaptive learning is not clear. Two related questions arise from the review of computational models. First, do uncertainty computations rely on estimates of stimulus or action values that are separate from those used to make decisions? Second, is expected uncertainty used as a baseline against which unexpected outcomes or events can be detected? As reviewed here, there is evidence that both the ACC⁴⁵ and OFC⁶⁸, and perhaps also the hippocampus, encode expected uncertainty. A critical follow-up question is whether this information is used to detect surprise and to update learning in subcortical regions. These questions can be addressed by independently estimating stimulus or action values and perceived uncertainty via subjective report and by manipulations of activity in brain areas encoding and computing those variables.

How does uncertainty affect value updating?

To allow effective learning, the brain must achieve a balance of scaling down learning when expected uncertainty is high versus scaling up learning when unexpected uncertainty is high. Therefore, understanding interactions between expected and unexpected uncertainty is crucial for understanding learning in naturalistic settings. However, it is unclear how the updating of stimulus or action values in the striatum, ACC and OFC depends on expected and unexpected uncertainty signals from the BLA and MD. Do expected and unexpected uncertainty signals directly influence the gain of learning, or are they instead read by an intermediate brain area? It is also unclear whether and how surprising events and unexpected uncertainty scale the gain of learning and how expected uncertainty scales down the gain of learning. These questions can be investigated via manipulations of neuromodulatory systems involved in learning.

Are reward prediction errors signals used to compute uncertainty?

Signed and unsigned reward prediction error signals are found in many brain areas, but exactly how they are used to compute expected uncertainty is unclear. This can be tested via optogenetic manipulations of dopamine signalling and measuring the effects on expected uncertainty signals in cortical regions.

Is uncertainty generalized across stimuli or actions (and if so, how)?

Expected and unexpected uncertainty could be generalized across stimuli and/or sets of actions in the reward environment. If so, is expected uncertainty used to weight volatility associated with each stimulus or set of actions to compute the global volatility or unexpected uncertainty of an environment? These questions can be investigated via behavioural manipulations in more complex experimental paradigms (BOX 3).

The importance of the non-human primate studies, specifically in macaques, in this field cannot be overstated; it is the crucial link to understanding how the human brain copes with and learns under uncertainty. We believe that the lack of behavioural paradigms and accompanying rodent models that are inspired by studies in humans and monkeys may have slowed progress in understanding the causal, systems-level neural mechanisms that support such adaptive learning and choice under different forms of uncertainty. Part of this is a methodological issue: the circuit dissection technology is more advanced in rodents, but the behavioural paradigms rarely are designed to mimic those used in primates. Moreover, custom and novel tasks

are needed to systematically examine the nuances of interactions between expected and unexpected uncertainty (BOX 3).

Value-based learning is assumed to happen at the synaptic level, whereas interactions between expected and unexpected uncertainty rely on circuit-level mechanisms. Therefore, revealing mechanisms of learning under uncertainty requires an understanding of the interactions between neural elements across multiple levels (synaptic and circuit-level), which is not possible without detailed computational modelling. Such models are also instrumental to computational psychiatry: various psychiatric conditions (including behavioural and substance addictions and anxiety disorders) are associated with either

failures in generating accurate models of the reward environment¹⁰⁴ or an inability to use those models to flexibly guide behaviour¹⁰⁵.

We hope that this Opinion outlines some important considerations for identifying the basic underlying mechanisms that may go awry in several neuropsychiatric disorders. Ultimately, a combination of novel behavioural paradigms, detailed mechanistic models, multi-area recording and circuit-level manipulations is required to answer critical lingering questions about learning under uncertainty (BOX 4).

Alireza Soltani¹ and Alicia Izquierdo²*

¹Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA.

²Department of Psychology, The Brain Research Institute, University of California, Los Angeles, Los Angeles, CA, USA.

*e-mail: soltani@dartmouth.edu; aizquie@psych.ucla.edu

<https://doi.org/10.1038/s41583-019-0180-y>

Published online: 30 May 2019

1. Doya, K. Modulators of decision making. *Nat. Neurosci.* **11**, 410–416 (2008).
2. Farashahi, S. et al. Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron* **94**, 401–414 (2017).
3. Iigaya, K. Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. *eLife* **5**, e18073 (2016).
4. Khorsand, P. & Soltani, A. Optimal structure of metaplasticity for adaptive learning. *PLoS Comput. Biol.* **13**, e1005630 (2017).
5. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat. Neurosci.* **3**, S1218–S1223 (2000).
6. Courville, A. C., Daw, N. D. & Touretzky, D. S. Bayesian theories of conditioning in a changing world. *Trends Cogn. Sci.* **10**, 294–300 (2006).
7. Bach, D. R. & Dolan, R. J. Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13**, 572–586 (2012).
8. McDannald, M. A. et al. Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. *Eur. J. Neurosci.* **35**, 991–996 (2012).
9. Langdon, A. J. et al. Model-based predictions for dopamine. *Curr. Opin. Neurobiol.* **49**, 1–7 (2018).
10. Tobler, P. N. et al. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J. Neurophysiol.* **97**, 1621–1632 (2007).
11. O'Reilly, J. X. Making predictions in a changing world: inference, uncertainty, and learning. *Front. Neurosci.* **7**, 105 (2013).
12. Preusschoff, K. & Bossaerts, P. Adding prediction risk to the theory of reward learning. *Ann. NY Acad. Sci.* **1104**, 135–146 (2007).
13. Diederer, K. M. & Schultz, W. Scaling prediction errors to reward variability benefits error-driven learning in humans. *J. Neurophysiol.* **114**, 1628–1640 (2015).
14. Yu, A. J. & Dayan, P. Uncertainty, neuromodulation, and attention. *Neuron* **46**, 681–692 (2005).
15. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput. Biol.* **7**, e1001048 (2011).
16. Faraji, M., Preusschoff, K. & Gerstner, W. Balancing new against old information: the role of puzzlement surprise in learning. *Neural Comput.* **30**, 34–83 (2018).
17. Jang, A. I. et al. The role of frontal cortical and medial-temporal lobe brain areas in learning a Bayesian prior belief on reversals. *J. Neurosci.* **35**, 11751–11760 (2015).
18. Chen, W. J. & Krajbich, I. Computational modeling of epiphany learning. *Proc. Natl Acad. Sci. USA* **114**, 4637–4642 (2017).

19. Nassar, M. R. et al. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* **30**, 12366–12378 (2010).
20. Behrens, T. E. et al. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
21. Costa, V. D. et al. Reversal learning and dopamine: a Bayesian perspective. *J. Neurosci.* **35**, 2407–2416 (2015).
22. Mathys, C. et al. A Bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39 (2011).
23. Funamizu, A. et al. Uncertainty in action-value estimation affects both action choice and learning rate of the choice behaviors of rats. *Eur. J. Neurosci.* **35**, 1180–1189 (2012).
24. Payzan-LeNestour, E. et al. The neural representation of unexpected uncertainty during value-based decision making. *Neuron* **79**, 191–201 (2013).
25. Wilson, R. C., Nassar, M. R. & Gold, J. I. A mixture of delta-rules approximation to Bayesian inference in change-point problems. *PLOS Comput. Biol.* **9**, e1003150 (2013).
26. McGuire, J. T. et al. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870–881 (2014).
27. Sutton, R. S. & Barto, A. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
28. Pearce, J. M. & Hall, G. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87**, 532–552 (1980).
29. Roesch, M. R. et al. Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. *Eur. J. Neurosci.* **35**, 1190–1200 (2012).
30. Krugel, L. K. et al. Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc. Natl Acad. Sci. USA* **106**, 17951–17956 (2009).
31. Dayan, P. & Long, T. In *Advances in Neural Information Processing Systems 10: Proceedings of the 1997 Conference* (eds Jordan, M. I., Kearns, M. J. & Solla, S. A.) 117–123 (MIT Press, 1998).
32. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
33. Soltani, A. C., W. & Wang, X. J. In *Decision Neuroscience: An Integrative Perspective* (eds Dreher, J.-C. & Tremblay, L.) 163–222 (Elsevier Academic Press, 2017).
34. Berridge, K. C. & Robinson, T. E. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res. Brain Res. Rev.* **28**, 309–369 (1998).
35. Salamone, J. D. et al. Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Curr. Opin. Pharmacol.* **5**, 34–41 (2005).
36. Redgrave, P. & Gurney, K. The short-latency dopamine signal: a role in discovering novel actions? *Nat. Rev. Neurosci.* **7**, 967–975 (2006).
37. Abraham, W. C. Metaplasticity: tuning synapses and networks for plasticity. *Nat. Rev. Neurosci.* **9**, 387 (2008).
38. Walton, M. E. et al. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. *Neuron* **65**, 927–939 (2010).
39. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
40. Niv, Y., Duff, M. O. & Dayan, P. Dopamine, uncertainty and TD learning. *Behav. Brain Funct.* **1**, 6 (2005).
41. Gershman, S. J. Dopamine, inference, and uncertainty. *Neural Comput.* **29**, 3311–3326 (2017).
42. Rogers, R. D. The roles of dopamine and serotonin in decision making: evidence from pharmacological experiments in humans. *Neuropsychopharmacology* **36**, 114–132 (2011).
43. Rushworth, M. F. & Behrens, T. E. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* **11**, 389–397 (2008).
44. Hayden, B. Y. et al. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J. Neurosci.* **31**, 4178–4187 (2011).
45. Monosov, I. E. Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nat. Commun.* **8**, 134 (2017).
46. Seo, H. & Lee, D. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
47. Hyman, J. M., Holroyd, C. B. & Seamans, J. K. A. Novel neural prediction error found in anterior cingulate cortex ensembles. *Neuron* **95**, 447–456 (2017).
48. Amiez, C., Joseph, J. P. & Procyk, E. Anterior cingulate error-related activity is modulated by predicted reward. *Eur. J. Neurosci.* **21**, 3447–3452 (2005).
49. Sul, J. H. et al. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* **66**, 449–460 (2010).
50. Stalnaker, T. A. et al. Orbitofrontal neurons signal reward predictions, not reward prediction errors. *Neurobiol. Learn. Mem.* **153**, 137–143 (2018).
51. Stalnaker, T. A., Cooch, N. K. & Schoenbaum, G. What the orbitofrontal cortex does not do. *Nat. Neurosci.* **18**, 620–627 (2015).
52. Riceberg, J. S. & Shapiro, M. L. Orbitofrontal cortex signals expected outcomes with predictive codes when stable contingencies promote the integration of reward history. *J. Neurosci.* **37**, 2010–2021 (2017).
53. Jo, S. & Jung, M. W. Differential coding of uncertain reward in rat insular and orbitofrontal cortex. *Sci. Rep.* **6**, 24085 (2016).
54. Riceberg, J. S. & Shapiro, M. L. Reward stability determines the contribution of orbitofrontal cortex to adaptive behavior. *J. Neurosci.* **32**, 16402–16409 (2012).
55. Izquierdo, A. Functional heterogeneity within rat orbitofrontal cortex in reward learning and decision making. *J. Neurosci.* **37**, 10529–10540 (2017).
56. Wallis, J. D. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* **15**, 13–19 (2011).
57. Rich, E. L. & Wallis, J. D. Decoding subjective decisions from orbitofrontal cortex. *Nat. Neurosci.* **19**, 973–980 (2016).
58. O'Neill, M. & Schultz, W. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. *Neuron* **68**, 789–800 (2010).
59. Saez, R. A. et al. Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* **95**, 70–77 (2017).
60. Massi, B., Donahue, C. H. & Lee, D. Volatility facilitates value updating in the prefrontal cortex. *Neuron* **99**, 598–608 (2018).
61. Paus, T. Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nat. Rev. Neurosci.* **2**, 417–424 (2001).
62. Heilbronner, S. R. & Hayden, B. Y. Dorsal anterior cingulate cortex: a bottom-up view. *Annu. Rev. Neurosci.* **39**, 149–170 (2016).
63. Rushworth, M. F. et al. Frontal cortex and reward-guided learning and decision-making. *Neuron* **70**, 1054–1069 (2011).
64. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013).
65. Kennerley, S. W. et al. Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* **9**, 940–947 (2006).
66. Winstanley, C. A. & Floresco, S. B. Deciphering decision making: variation in animal models of effort- and uncertainty-based choice reveals distinct neural circuitries underlying core cognitive processes. *J. Neurosci.* **36**, 12069–12079 (2016).
67. Mobini, S. et al. Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probabilistic reinforcement. *Psychopharmacology* **160**, 290–298 (2002).
68. Stolyarova, A. & Izquierdo, A. Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. *eLife* **6**, e27483 (2017).
69. Dalton, G. L. et al. Multifaceted contributions by different regions of the orbitofrontal and medial prefrontal cortex to probabilistic reversal learning. *J. Neurosci.* **36**, 1996–2006 (2016).
70. Bradfield, L. A. et al. Medial orbitofrontal cortex mediates outcome retrieval in partially observable task situations. *Neuron* **88**, 1268–1280 (2015).
71. Rudebeck, P. H. et al. Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes. *Neuron* **95**, 1208–1220 (2017).
72. Noonan, M. P. et al. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl Acad. Sci. USA* **107**, 20547–20552 (2010).
73. Meder, D. et al. Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. *Nat. Commun.* **8**, 1942 (2017).
74. Heilbronner, S. R. et al. Circuit-based corticostriatal homologies between rat and primate. *Biol. Psychiatry* **80**, 509–521 (2016).
75. Vogt, B. A. & Paxinos, G. Cytoarchitecture of mouse and rat cingulate cortex with human homologies. *Brain Struct. Funct.* **219**, 185–192 (2014).
76. Hoover, W. B. & Vertes, R. P. Projections of the medial orbital and ventral orbital cortex in the rat. *J. Comp. Neurol.* **519**, 3766–3801 (2011).
77. Hunt, L. T. et al. Triple dissociation of attention and decision computations across prefrontal cortex. *Nat. Neurosci.* **21**, 1471–1481 (2018).
78. White, J. K. & Monosov, I. E. Neurons in the primate dorsal striatum signal the uncertainty of object-reward associations. *Nat. Commun.* **7**, 12735 (2016).
79. Costa, V. D. et al. Amygdala and ventral striatum make distinct contributions to reinforcement learning. *Neuron* **92**, 505–517 (2016).
80. St Onge, J. R. et al. Separate prefrontal-subcortical circuits mediate different components of risk-based decision making. *J. Neurosci.* **32**, 2886–2899 (2012).
81. Aeverbeck, B. B. & Costa, V. D. Motivational neural circuits underlying reinforcement learning. *Nat. Neurosci.* **20**, 505–512 (2017).
82. Monosov, I. E. & Hikosaka, O. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat. Neurosci.* **16**, 756–762 (2013).
83. Unal, G. et al. Synaptic targets of medial septal projections in the hippocampus and extrahippocampal cortices of the mouse. *J. Neurosci.* **35**, 15812–15826 (2015).
84. Kumaran, D. & Maguire, E. A. An unexpected sequence of events: mismatch detection in the human hippocampus. *PLOS Biol.* **4**, e424 (2006).
85. Vanni-Mercier, G. et al. The hippocampus codes the uncertainty of cue-outcome associations: an intracranial electrophysiological study in humans. *J. Neurosci.* **29**, 5287–5294 (2009).
86. Wikenheiser, A. M. & Schoenbaum, G. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nat. Rev. Neurosci.* **17**, 513–523 (2016).
87. Wikenheiser, A. M. & Redish, A. D. Decoding the cognitive map: ensemble hippocampal sequences and decision making. *Curr. Opin. Neurobiol.* **32**, 8–15 (2015).
88. Morrison, S. E. et al. Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* **71**, 1127–1140 (2011).
89. Rudebeck, P. H. et al. Amygdala contributions to stimulus-reward encoding in the macaque medial and orbital frontal cortex during learning. *J. Neurosci.* **37**, 2186–2202 (2017).
90. Saez, A. et al. Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* **87**, 869–881 (2015).
91. Wassum, K. M. & Izquierdo, A. The basolateral amygdala in reward learning and addition. *Neurosci. Biobehav. Rev.* **57**, 271–283 (2015).
92. Roesch, M. R. et al. Neural correlates of variations in event processing during learning in basolateral amygdala. *J. Neurosci.* **30**, 2464–2471 (2010).
93. Cassell, M. D. & Wright, D. J. Topography of projections from the medial prefrontal cortex to the amygdala in the rat. *Brain Res. Bull.* **17**, 321–333 (1986).
94. Amaral, D. G. & Price, J. L. Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *J. Comp. Neurol.* **230**, 465–496 (1984).
95. Sharpe, M. J. & Schoenbaum, G. Back to basics: making predictions in the orbitofrontal-amygdala circuit. *Neurobiol. Learn. Mem.* **131**, 201–206 (2016).
96. Lucantonio, F. et al. Neural estimates of imagined outcomes in basolateral amygdala depend on orbitofrontal cortex. *J. Neurosci.* **35**, 16521–16530 (2015).
97. Stopper, C. M. et al. Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron* **84**, 177–189 (2014).
98. Mitchell, A. S., Baxter, M. G. & Gaffan, D. Dissociable performance on scene learning and strategy implementation after lesions to magnocellular mediodorsal thalamic nucleus. *J. Neurosci.* **27**, 11888–11895 (2007).
99. Izquierdo, A. & Murray, E. A. Functional interaction of medial mediodorsal thalamic nucleus but not nucleus accumbens with amygdala and orbital prefrontal cortex is essential for adaptive response selection after reinforcer devaluation. *J. Neurosci.* **30**, 661–669 (2010).

100. Mitchell, A. S. et al. Advances in understanding mechanisms of thalamic relays in cognition and behavior. *J. Neurosci.* **34**, 15340–15346 (2014).
101. Chakraborty, S. et al. Critical role for the mediodorsal thalamus in permitting rapid reward-guided updating in stochastic reward environments. *eLife* **5**, e13588 (2016).
102. Parnaudeau, S. et al. Mediodorsal thalamus hypofunction impairs flexible goal-directed behavior. *Biol. Psychiatry* **77**, 445–453 (2015).
103. Wolff, M. & Vann, S. D. The cognitive thalamus as a gateway to mental representations. *J. Neurosci.* **39**, 3–14 (2019).
104. Voon, V. et al. Model-based control in dimensional psychiatry. *Biol. Psychiatry* **82**, 391–400 (2017).
105. Vaghi, M. M. et al. Compulsivity reveals a novel dissociation between action and confidence. *Neuron* **96**, 348–354 (2017).
106. Soltani, A. & Wang, X. J. A biophysically based neural model of matching law behavior: melioration by stochastic synapses. *J. Neurosci.* **26**, 3731–3744 (2006).
107. Soltani, A. & Wang, X. J. From biophysics to cognition: reward-dependent adaptive choice behavior. *Curr. Opin. Neurobiol.* **18**, 209–216 (2008).
108. Izquierdo, A. et al. The neural basis of reversal learning: an updated perspective. *Neuroscience* **345**, 12–26 (2017).
109. Cardinal, R. N. Neural systems implicated in delayed and probabilistic reinforcement. *Neural Netw.* **19**, 1277–1301 (2006).
110. Cardinal, R. N. & Howes, N. J. Effects of lesions of the nucleus accumbens core on choice between small certain rewards and large uncertain rewards in rats. *BMC Neurosci.* **6**, 37 (2005).
111. Ghods-Sharifi, S., St Onge, J. R. & Floresco, S. B. Fundamental contribution by the basolateral amygdala to different forms of decision making. *J. Neurosci.* **29**, 5251–5259 (2009).
112. Li, Y. & Dudman, J. T. Mice infer probabilistic models for timing. *Proc. Natl Acad. Sci. USA* **110**, 17154–17159 (2013).
113. Dalton, G. L., Phillips, A. G. & Floresco, S. B. Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts. *J. Neurosci.* **34**, 4618–4626 (2014).
114. Donahue, C. H. & Lee, D. Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nat. Neurosci.* **18**, 295–301 (2015).
115. Amodeo, L. R., McMurray, M. S. & Roitman, J. D. Orbitofrontal cortex reflects changes in response-outcome contingencies during probabilistic reversal learning. *Neuroscience* **345**, 27–37 (2017).
116. Daw, N. D. et al. Cortical substrates for exploratory decisions in humans. *Nature* **441**, 876–879 (2006).
117. Averbeck, B. B. Theory of choice in bandit, information sampling and foraging tasks. *PLOS Comput. Biol.* **11**, e1004164 (2015).
118. Groman, S. M. et al. Chronic exposure to methamphetamine disrupts reinforcement-based decision making in rats. *Neuropsychopharmacology* **43**, 770–780 (2018).
119. Groman, S. M. et al. Dopamine D3 receptor availability is associated with inflexible decision making. *J. Neurosci.* **36**, 6732–6741 (2016).

Acknowledgements

The authors thank D. Lee, P. Rudebeck and A. Wikenheiser for helpful feedback. The authors acknowledge support from the US National Institutes of Health Grant R01DA047870 (A.S. and A.I.), a University of California–Los Angeles (UCLA) Division of Life Sciences Recruitment and Retention Fund (A.I.) and a UCLA Academic Senate Grant (A.I.).

Author contributions

Both authors researched data for the article, made substantial contributions to discussion of the content, wrote the article and reviewed or edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reviewer information

Nature Reviews Neuroscience thanks S. B. Floresco and K. Preuschoff, and the other anonymous reviewer(s), for their contribution to the peer review of this work.