How outcome uncertainty mediates attention, learning, and decision-making

Ilya E. Monosov [Associate Professor, Adjunct Professor, Principal Investigator]

Departments of Neuroscience and Neurosurgery, Department of Biomedical Engineering, Washington University Pain Center, Washington University School of Medicine in St. Louis

Abstract

Animals and humans evolved sophisticated nervous systems which endowed them with the ability to form internal-models or beliefs, and make predictions about the future to survive and flourish in a world in which future outcomes are often uncertain. Crucial to this capacity is the ability to adjust behavioral and learning policies in response to the level of uncertainty. Until recently, the neuronal mechanisms that could underlie such uncertainty-guided control have been largely unknown. In this review, I discuss newly discovered neuronal circuits in primates that represent uncertainty about future rewards, and propose how they guide information-seeking, attention, decision-making, and learning to help us survive in an uncertain world. Lastly, I discuss the possible relevance of these findings to learning in artificial systems.

Keywords

Slow reproduction, high capacity for prediction: the mammalian central nervous system evolved to negotiate uncertainty

Surviving and taking advantage of unpredictable events is a fundamental challenge for all organisms [1-6]. At the population level, unexpected events and variable environments or contexts may drive genetic variability, population abundance, spatial dispersal, and alter swarm or group behavior [3, 4, 6, 7]. For example, unexpected stressors may lead single cell organisms to increase their rates of mutation and division [7-9]. Many such population level adaptation mechanisms function on rapid time scales and across many generations.

In animals that possess sophisticated nervous systems, unexpected events and variable contexts also activate behavioral adaptation mechanisms such as changes in foraging, action planning, motivational drive, valuation strategy, and the rate of learning new information, including rules and higher order statistical models of the environment [1, 6, 10-12].

Correspondence: ilya.monosov@gmail.com [I.E. Monosov].

Also, in their everyday decision making, many animals will forego some immediate amount of reward (or pleasure) for advance information that will reduce their uncertainty about difficult to predict events of high importance (such as rewards), even if that information cannot be used to modify their future [13].

The capacity for uncertainty-related behavioral adjustments is particularly crucial for animals such as mammals, and particularly for primates, due to our slow rate of replication, maturation, and our relatively slow population-level responses to variability or challenge.

In this review, I outline some of the key behavioral changes that primates employ to survive in an uncertain world and propose biologically plausible mechanisms to support them. The article primarily reviews recent electrophysiological studies in behaving monkeys and concerns itself with how explicit representations of uncertainty in the primate brain guide cognitive processing. The intent is that it, alongside other reviews that outline how prediction errors mediate our beliefs under uncertainty and painstakingly define distinct forms of uncertainty that animals and humans face [1, 11, 14-17], will guide future experiments to reveal the mechanisms of learning and adaptation in an uncertain world.

Behavioral goals change as a function of outcome uncertainty: an overview

Uncertainty about future rewards arises from many sources. Some rewards are uncertain because they are fundamentally probabilistic or *risky* (e.g., coin toss). However, rewards can also become uncertain due to many additional factors, including uncertainty about incoming sensory percepts and outgoing motor commands, the values of objects and actions, and the states and dynamics of the environment [1].

When rewards are available and uncertainty about them is low, subjects may increase the rate and speed of reward-driven behaviors [18-20] and aim to reduce the mental cost of those behaviors [20-22] so that those resources could support other ongoing cognitive processes. Automatic behaviors, such as skills or habits [18, 19, 23-25] which are efficient and require little online evaluation [26], are well-suited to obtain rewards quickly in low uncertainty environments and allow subjects to engage in additional goal directed behaviors. Also, they can be learned through classical reinforcement learning (RL) mechanisms in which if a reward gained is better than expected, the learning mechanism assigns credit for the benefit to the action(s) that preceded it, facilitating them to increase their probability or magnitude [19]. Under low uncertainty, this learning rule work exceedingly well.

When uncertainty about future rewards is high, the goals of learning often change [1, 27-29]. For instance, once the agent has detected high outcome uncertainty, particularly because the values of its action are changing, the performance of RL learning rules can be improved by incorporating additional meta-learning parameters based on estimates of uncertainty [30]. Agents can learn much faster if they adapt their learning rates to match the environment [31] (but see [12]). They can use a fast rate in some novel and volatile environments where a single large prediction error warrants a large behavioral change, or a slow learning rate in stable environments that change gradually, and in noisy environments where large prediction

errors are common but may not warrant changing behavior. The agent can accomplish these adjustments by using an internal estimate of reward uncertainty (Figure 1).

In high uncertainty, instead of seeking immediate reinforcement, actors must try to learn the structure of the world and reduce their uncertainty - otherwise, proper credit assignment may not be achieved, and learning-related behavioral enhancements may not actually result in *more* reward. Indeed, humans and animals assign value to obtaining information to resolve uncertainty about future rewards, even when this information cannot be used to directly and immediately to influence the timing or amount of reward they will get [32-37]. This may ultimately be beneficial in natural environments where accurate estimation and reduction of uncertainty can accelerate learning of adaptive behaviors in order to succeed in the future [38], and during foraging under threat, predation, or competition [39, 40].

To support uncertainty-related behavioral adaptation, when uncertainty is high, it may be necessary to lengthen the time-scale of learning [29], such that entire sequences of behaviorally relevant events are considered, which may require episodic memory, to facilitate processes such as transfer and one-shot learning. Agents may place special emphasis on "learning-to-learn" [41], perhaps partly through synaptic changes that mediate how neural circuits respond to surprising events [31] or by changing their behavioral strategy, such as for example through the facilitation of distinct modes of "attention" [16, 25, 42].

On the surface, it may seem that these uncertainty-related modulations should only be implemented if the agent thinks they are in a volatile or ambiguous environment, because there learning from outcomes and tracking uncertainty could theoretically decrease uncertainty (and increase confidence). However, it is likely that the same or similar processes take place under *risk.* I return to this point later in Concluding Remarks.

In summary, under reward uncertainty, behavior must be mediated such that an individual agent can resolve it, learn uncertainty, and obtain an internal representation of possible outcomes. This may occur by adjusting risk attitude (to change the rate of surprises), learning strategies, and by sacrificing physical or "cognitive" reward to gain advance information about the future. Many cognitive functions such as episodic memory and attention must sub serve these goals, and support the ability of the agent to select appropriate automatic and goal-oriented behaviors (and learning strategies that support them).

Next, I will discuss the circuits that represent uncertainty about future rewards and discuss the possible advantages of explicitly representing uncertainty for behavioral and cognitive control.

Multiple neuronal representations of reward uncertainty in the primate brain

Does the brain encode reward uncertainty with a distributed population code? Or does it also contain a representation of uncertainty that can be read out from the activity of single neurons? Neuroscience has made strides in answering similar questions in the sensory

domain. When it is due to variability in sensory inputs, or expressed in the form of uncertainty regarding action output, uncertainty is represented in the population dynamics of neuronal pools [43].

Differentially, recent work also showed that primate brain regions involved in value-based decision making and reinforcement learning contain single neurons that vary their discharge rates with quantitative levels of reward uncertainty. These single reward uncertainty selective neurons are located in two networks: in the septal and basal forebrain (BF) areas in the center of the forebrain, and in an interconnected network of cortico-basal ganglia regions that include the anterior cingulate cortex, the internal capsule bordering striatum, and the ventral regions of the anterior pallidum [33, 44-48], which I will refer to as the ACC-icbDS-Pallidum network.

Also, one study found that the orbitofrontal cortex (OFC) contains neurons that scale with reward uncertainty [49]. There, as the standard deviation of rewards increased, the monkeys' preference also increased. At the same time, the neuronal activity of some OFC neurons was correlated with the monkeys' preferences for higher reward variance. Importantly, the same neurons' activity did not scale with reward magnitude. This result demonstrates that embedded within the OFC value- decision related circuits there are a minority of neurons that process information about reward uncertainty.

Why does the brain contain multiple circuits and areas that signal reward uncertainty? To answer the question, I will discuss their differences and concentrate on the BF and ACC-icbDS-Pallidum networks, whose reward uncertainty neurons have been particularly well characterized. I will then go on to argue that the BF is particularly well-suited to play a global coordinative function, mediating cortical activity across many brain areas to facilitate learning and attention under uncertainty, whereas the ACC-icbDS-Pallidum network particularly motivates behaviors aiming to obtain uncertainty-resolving information.

Both BF and ACC-icbDS-Pallidum uncertainty neurons on average have uncertainty related "ramp-like" activity: that is, following an uncertain prediction, the average activity of many single neurons increased (or decreased) as a function of time until the uncertainty was resolved [33, 44, 47, 50]. However, the ramping in the BF and ACC-icbDS-Pallidum differed in several important ways. BF reward uncertainty neurons' ramping was not strongly dependent on the presence of the reward uncertainty predicting visual stimuli and was not highly sensitive to gaze parameters. They even displayed ramping activity when animals' eyes were closed during blinking [45, 47] and during trace conditioning in which no stimuli were present on the display [44, 47]. In contrast, the ACC-icbDS-Pallidum network was highly sensitive to the presence of objects associated with reward uncertainty. The network's signals were greatly *decreased* when uncertain objects were removed [47], most crucially the network's signals *increased* before monkeys made gaze shifts to uncertain visual objects that predicted advance information that resolved the uncertainty [33].

Another difference between BF and ACC-icbDS-Pallidum uncertainty neurons emerged when we tested whether and how uncertainty-selective neurons dynamically updated their representation of uncertainty as animals learned novel probabilistic object reward

associations. At the start of learning, the ACC-icbDS-Pallidum network strongly responded to all 3 visual cues. But, within the same session, as the monkeys learned the objects' reward-predictions, the networks' uncertainty signals decreased for the certain reward associated images (0 and 100%) [47]. In contrast, BF uncertainty ramping activity, as well as saccadic choice response times, changed gradually across multiple days [45], as the monkeys gained familiarity with the images and became fast and skillful at utilizing them to harness rewards [51]. Finally, almost all BF uncertainty sensitive neurons ramp to the time of certain noxious events and to the expected time of presentations of novel objects [44, 45]. So, relatively, BF ramping is a more general expectation mechanism.

With these differences in mind, next I will discuss how the less specific BF and the relatively more specific ACC-icbDS-Pallidal reward uncertainty signals could mediate attention, learning, and decision making under uncertainty.

Memory formation and credit assignment under reward uncertainty

Animals face a unique set of challenges in naturalistic environments in which many types of uncertainties co-occur. As a result, they are faced with a credit assignment problem: they must decide which previous cues and actions predicted or caused rewards or punishments. This is no easy feat, as in natural environments, the influences of sensory noise and ambiguity, uncertainty about the structure of the world, and outcome uncertainty are compounded, making it difficult to decide which events in the past led to which outcomes in the present. To aid us in solving this credit assignment problem, the brain may have adopted strategies that rely on memories and attentional control (e.g., to help us gain a model of the environment). In support of these processes, uncertainty could promote memory formation by explicitly changing the motivation to learn, by mediating plasticity [52, 53], and by influencing attention including overt and covert spatial attention [16, 33, 54].

A recent behavioral study in humans has shown that recognition memory enhancements in reward variable contexts may occur not only due to the reward prediction errors that generally occur after uncertainty or risk is resolved [55]. That is, when high reward variance was associated with a particular object, recognition memory reports about that object were more accurate. Such memory enhancements under reward uncertainty may be influenced by attentional and (meta)plastic mechanisms. Objects experienced under uncertainty may be preferentially processed due to covert or overt attention resulting in a richer or finer sensory representation leading to enhanced object or context memory. At the same time, uncertainty may also increase the capacity of synapses to learn and store information. Theoretical and behavioral work [15, 52] argues that meta plasticity could contribute to the estimation of uncertainty. But, whether neuronal circuits that represent uncertainty mediate meta plasticity remains an open question.

An adaptive way for the brain to regulate memory formation and attentional processes could be by encoding predictions about the expected time of upcoming surprising, important, rewarding, and noxious outcomes. That is because events that resolve uncertainty or that are proximal to behaviorally salient outcomes are particularly important to commit to memory. Such neural predictions could take the form of population ramping activity, in which case

they would regulate the capacity to learn as a function of time – increasing learning and memory in anticipation of the time of the upcoming outcome, and peaking as the outcome itself arrives (and the need for learning and memory encoding is greatest).

Next, I will briefly describe the circuitry that could give rise to such a general surprise expectation signal, and then will argue that the BF uncertainty-related ramping neurons are a likely candidate to broadcast this signal to the cortical mantle.

The origin and functional role of the basal forebrain ramping activity

Recent evidence suggests several cortical and subcortical brain areas could support distinct aspects of such temporal predictions that could ultimately be integrated to give rise to ramping activity that anticipates a wide range of behaviorally salient events. For example, the basal ganglia contain neurons that could underlie population-level encoding of time through coordinated phasic activity [56]. Also, the dorsal medial prefrontal cortex and the hippocampus [57, 58] possess context specific temporal codes. Finally, the ACC contains valence-specific neurons, distinctly predicting the timing of rewards or punishments [48] (see Box 1). In principle, this diversity could facilitate the prediction of many distinct events for the control of a diverse set of action plans, emotional states, and memories.

The context-dependent temporal codes in these brain areas may contribute to relatively non-specific ramping activity in the BF [56]. And, the resulting signal may be particularly well suited to signal temporal predictions for the purpose of enhancing memory formation and mediating a wide range of sensory and cognitive processes (Figure 2). First, BF ramping anticipates uncertain outcomes as well as many other kinds of behaviorally important events [44, 45, 50]. Second, BF ramping is relatively independent of sensory or motor state. Third, BF ramping is strongly linked to beliefs about the timing and structure of the behavioral context [44]. Fourth, BF is anatomically well-positioned to coordinate and simultaneously change the gain [59] of a wide range of neuronal circuits (Figure 2A-C) because it projects to the entire cortical mantle [60] – which would be crucial for mnemonic and attentional functions that rely on wide coordination across distinct functional regions of cortex.

One particularly salient anatomical observation is that in primates the BF strongly projects to the medial temporal lobe [45, 60]. This BF-temporal circuit, particularly the cholinergic input from the BF, is necessary for object memory formation [61]. As discussed earlier, BF uncertainty-sensitive ramping neurons' activity changes slowly during new object value learning and may depend on object familiarity (and not purely on the explicit knowledge of object values [45]). Such slow learning rates are reflected in monkeys' gaze behavior [18, 45, 51]. One possibility is that these slow changes in the saccadic behavior could be directly mediated by the BF-temporal cortex circuit. Under the influence of the BF, the temporal cortex would then mediate gaze through the posterior ventral basal ganglia, which learns object salience over long time scales, receives dense projections from most areas of the temporal cortex, and directly controls the generation of eye movements [62].

The BF is also well positioned to control cortical computations under uncertainty through a distinct population of phasically activated neurons that rapidly signal the occurrence of

behaviorally salient external events (Figure 2B). These neurons are sensitive to the degree of statistical errors, for example in reward value or in sensory patterns like in object sequences, display strong bursting activations in responses to novel objects [44, 63, 64].

In summary, ramping BF neurons may coordinate learning, memory, and attention by mediating the gain and spike timing of many cortical regions (Figure 2C) relative to the agent's predictions and expectations. And, in concert, the precise short latency BF phasic bursting activity could signal the occurrence of external behaviorally salient and uncertain events in a temporally precise manner. BF ramping activity may increase or decrease the effect of this widely broadcasted phasic salience signal, but how it would do so remains unknown. Broadly consistent with the proposed coordinative functions of the BF, inactivation of the BF changes the correlation structure of primate cortical brain areas that receive the densest BF inputs from the inactivated/disrupted BF region [65]. And, BF inputs to the cortical mantle directly impact global states [66] in manners that could facilitate learning and attention [67]. Finally, modelling and experimental data indicate that cholinergic manipulations directly impact attention-related computations in multiple brain areas, including the frontal eye field, a key region for the mediation of spatial selective attention [68-73].

To date, how BF inputs accomplish this remains unclear, particularly in primates. Therefore, to test the mechanisms proposed in Figure 2C, future work must assess the relative contribution of the phasic and ramping cell types to cortical computations by (1) specifically understanding which cortical neurons receive their inputs and (2) which neurotransmitter(s) they release.

Reward uncertainty mediates information seeking gaze behavior: the role of the ACC-icbDS-Pallidal network

Visual objects that predict uncertain outcomes are often associated with a chance to learn and update our beliefs because they are followed by prediction errors. Therefore, many theories predict that uncertain objects ought to attract gaze because gaze is the chief instrument of primate information seeking.

However, thus far most experiments showed that the speed and probability of saccadic gaze shifts to visual objects are strongly driven by their absolute values or behavioral salience [51, 74]: objects associated with highly probable big rewards or with noxious, aversive, or intense outcomes preferentially attract gaze [48, 74]. This is particularly the case immediately after object presentation [48, 74].

In recent years a considerable effort has been undertaken to also understand how uncertainty impacts primate gaze behavior. One study showed that when information was available and could be obtained with saccades, saccadic choices and response times reflected the motivational value of information to resolve reward uncertainty [32]. A later study replicated and extended this important observation to visual search, showing that animals examined more search targets, and did so faster or for longer durations, to gather uncertainty-resolving information [75]. Moreover, we subsequently found that even at times when information was

not yet available and was not contingent on gaze, animals still preferentially gazed at objects associated with uncertain rewards, anticipating the moment when information would arrive [48]. These results show that primates are motivated to resolve uncertainty and assign high motivational value to objects that predict information gain (uncertainty reduction). Their desire to seek information is robustly reflected in gaze patterns.

Information seeking may be in part supported by reward circuitry in manners that overlap with reward seeking. The phasic bursting of dopamine neurons signaled the possibility to resolve uncertainty earlier in time similarly to how they signaled fluid rewards [34]. However, a key question remained: Through what mechanism might the motivational drive to seek advance information be implemented? While phasic dopamine neuron responses could be used during learning to endow informative cues with value, it was unknown how neural circuits translated those values into motivation to guide online information seeking behavior.

To answer this question, we designed an *information viewing* procedure that studied whether monkeys naturally exhibit gaze behavior that reflects their anticipation of uncertainty-resolving information. In this procedure, there was no way for the monkeys to use information or their gaze patterns to influence reward magnitude. We then studied the relationship of the monkeys' uninstructed gaze behavior and the neural activity in the ACC-icbDS-pallidal network (Figure 3).

Behaviorally, we found that the monkeys' gaze was strongly attracted to objects based on their uncertainty, especially in the moments before receiving information to resolve it. This attraction was specifically related to anticipating information (Figure 3B), not reward value or uncertainty per se [33].

Because spatial attention can influence both gaze and visual processing, and gaze behavior reflects attentional biases, it is possible that this object importance signal influences visual processing itself. So, when primates first encounter a visual object (perhaps during the initial low-level feature analyses of that object), its relative importance (based on stored memory) has a great impact on gaze. This importance is directly related to many factors such as reward, punishment, uncertainty, and information. Next, after the gaze is on the object or after it has been recognized, reward uncertainty of the object has a particularly powerful effect on gaze, especially before the time of uncertainty resolution (Figure 3B).

Our work suggested that this attraction may be directly related to, and serve, information seeking, and that information seeking is regulated by the ACC-icbDS-pallidal network [33]. The data indicate that single neurons in the ACC-icbDS-pallidal network (A) monitor the level of uncertainty about future events (B) anticipate the timing of information, and (C) anticipate information-seeking actions, such as gaze shifts to inspect the source of uncertainty. The uncertainty sensitive neurons in the network predicted when information will become available in *information viewing procedure* (Figure 3C). Their activity was most prominent during uncertain trials and ramped up to times when the monkeys expected information: when they were anticipating an informative visual cue or when they had not received an informative cue and were instead anticipating being informed of the outcome by

reward delivery or omission itself at the end of the trial. Furthermore, this information-anticipatory activity was coupled to gaze behavior on a moment-by-moment basis: increases in this activity were followed by gaze shifts *to* uncertainty-resolving objects, and decreases in this activity were followed by gaze shifts *away* from uncertainty-resolving objects (Figure 3D).

Surprisingly, fluctuations in ACC information signals were the earliest predictor of future behavior. ACC information signals began to change hundreds of milliseconds before these gaze shifts, while the icbDS-pallidal regions of the network changed their activity more proximally to behavior. So, the ACC may be more directly linked to the cognitive processes surrounding the drive to seek advance information to resolve uncertainty, while the basal ganglia may be more directly related to the implementation of this drive to control action. Also, consistent with the role of the basal ganglia in action control, temporary chemical disruptions of icbDS-Pal activity disrupted monkeys' information seeking (Figure 3E). This finding not only provides direct evidence that the information predicting signals in the basal ganglia mediate information seeking, but also show that the anterior basal ganglia loops have functions beyond those predicted by classical reinforcement learning.

How would the ACC-icbDS-Pallidum network control gaze? One candidate mechanism is through the basal ganglia which influence saccades and attention by mediating the activity of the superior colliculus [76]. Though the anterior pallidal regions in the information seeking network may not send strong direct projections to the superior colliculus, the ACC-icbDS-Pallidum network is well-suite to directly impact other basal ganglia loops. For example, the dorsal caudate nucleus has been shown to carry an oculomotor signal that combines information about the spatial location and the reward uncertainty of upcoming saccadic targets [77]. Control of gaze by uncertainty may also take place through a cortical route. For example, LIP, a parietal area that contributes to saccadic behavior and attention, displays modulations of its gaze-related spatially selective activity by the subjects' motivation to resolve uncertainty [54]. Also, the frontal eye fields (FEF), another cortical area that directly impacts attention and gaze [71-73] and directly projects to dorsal caudate [78] and LIP, receives inputs from the ACC and could be another possible candidate site through which uncertainty signals impact information seeking related gaze behavior.

Based on these observations and our previous studies, Figure 3F proposes a circuit/anatomy constrained mechanism for information seeking. The motivation to seek uncertainty resolving information is mediated by the ACC-icbDS-Pallidum network. Anatomically, pallidum projects directly to dopamine neurons and also mediates dopamine through the lateral habenula – brainstem circuitry [79, 80]. Because the ACC-icbDS-Pallidum network anticipates the timing of uncertainty resolving prediction errors (Figure 3C), through this pathway, ramping activity can regulate dopaminergic prediction error related learning. Also, dopamine neurons project to ACC-icbDS-Pallidum network, particularly strongly to the striatum. Because dopamine neurons signal prediction errors that result from uncertainty resolution by reward-predicting information [32], for example following the time of the info cue in Figure 3C, dopaminergic signaling is perfectly suited to endow information with value through shaping of cortico-striatal interactions. Finally, to obtain information primates rely on gaze. Both dopamine and ACC-icbDS-Pallidum network can mediate information

seeking directly by controlling gaze either through the basal ganglia oculomotor circuit (comprising of the striatum and substantia nigra pars reticulata) or through cortical pathways to the superior colliculus. The information obtained by observing the world, can then guide reinforcement learning and information seeking. The BF (Figure 2) projects to all cortical sites in Figure 3F. And though it is not shown there, it may play a crucial coordinative role in mediating information seeking across this wide range of circuits.

It is also worth mentioning that in situations in which economic (or subjective value driven-) decision making guides behavior, uncertainty reduction and information seeking may also be impacted by changes in risk attitude that coincide with probability distortions, leading to optimistic or pessimistic evaluations of the future. We previously proposed this idea based on observations of single neurons' activity in the primate ventromedial prefrontal (vmPF) cortex [81]. In this brain region, anatomically segregated populations of single neurons reflected the optimistic and pessimistic integration of uncertainty and expected value, for example leading to what we termed as a *hopeful* or overly-positive weighting of possible future rewards [81]. This may promote or suppress risky choices, changing the frequency of encountering prediction errors and thus potentially leading to adaptations in learning processes and uncertainty reduction. More recently, similar proposals have been made based on observations of single dopamine neuron activity [82]. Dopamine neurons appear to be heterogenous in their activations: some evaluating predictions optimistically or pessimistically, in a spectrum. Through this heterogeneity they appear to signal prediction errors based on the distribution of expected rewards rather than the mean of expected rewards. Importantly, depending on how optimistic or pessimistic dopamine neurons interact with other brain regions (e.g., via their anatomical projection patterns to different striatal areas that receive inputs from vmPF cortex), they may differentially and flexibly impact valuation of risky options by changing responses to offers or reward feedback, to subsequently mediate risk attitudes and impact uncertainty reduction and learning.

Concluding remarks

The study of how agents learn from outcomes under uncertainty remains a crucial and growing field in neuroscience. I have outlined how explicit representations of uncertainty could be utilized to control attention, memory, and learning. Next, I outline some important future directions for advancing our understanding of behavioral control under uncertainty and provide their summary in the Outstanding Questions.

To negotiate complex and uncertain environments agents must be willing to put forth mental effort, patience, and persistence in the face of failures and setbacks. Also, they must be willing to take risks to generate prediction errors in order to learn, and go on to estimate and/or reduce their uncertainty. While a large body of literature exists on risk-attitude in economics and behavioral neuroscience, very little research has been conducted on how neural representations of uncertainty mediate the drive to gain information to resolve it [13, 32, 33, 54], and on the relationship between information seeking, risk attitude, and mental effort. Future work should carefully assess the interaction of these factors within individual subjects, and derive the mental algorithms that govern them. This effort may also facilitate

the improvement of artificial neural networks, which at present tend to perform poorly under uncertainty (Box 2).

Another important question to study is whether and how uncertainty mediates automatic versus controlled actions, and more broadly whether and how it participates in action selection. Under high uncertainty, humans may maintain a high level of automaticity for many behaviors that are or have been immediately rewarding. We rely on this capacity to support us as we learn, perform difficult tasks, ruminate, explore our memories, and seek a better understanding of the world. Recent work indicates that distinct circuits within the ventral basal ganglia influence "automatic" gaze shifts that are based on the long-term memory versus goal-directed gaze shifts that are based on dynamic estimates of object values [18]. Future research should ask whether and how uncertainty controls different regions of the basal ganglia to mediate automatic and goal-directed actions.

An important distinction should be made between the concepts of risk and ambiguity. Under risk, uncertain outcomes are drawn from a known probability distribution, about which theoretically *nothing more can be learned.* By contrast, under ambiguity, uncertain outcomes are drawn from an unknown probability distribution which may be learnable through new experiences [12, 47, 83, 84]. Adaptive behavior may differ, in principle between risky and ambiguous situations. But in natural environments, range, entropy, or variance of rewards may change at any time. So, even under risk, following outcomes, agents ought to update their state because otherwise they would not know when these changes occur. In general, risk without volatility and ambiguity is extremely rare in nature. Therefore, the distinction between risk and ambiguity may not be categorical. There are rarely totally stationary or stable probability distributions, and even when they exist, it is hard for agents to know that with certainty. Given these considerations, I would argue that rather than pursuing the dissociation of risk and ambiguity at the level of neuronal circuits, effort ought to be dedicated to understanding how learning strategies change with the agent's beliefs about how stable or volatile the environment is, and/or how much of the variability in outcomes is due to an inherent variance or measurement noise. For example, recent work indicates that the ACC contains outcome-related neurons that signal feedback in a manner that is sensitive to the sources of uncertainty [85, 86]. Hence, along this line, an important direction for future work is find out how these uncertainty-resolution / post-decisional signals differentiate distinct sources of reward uncertainty that preceded them to adaptively mediate learning and attention.

Acknowledgements

References

1. Bach DR, and Dolan RJ (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. Nature reviews. Neuroscience 13, 572–586. [PubMed: 22781958]

2. Nefti S, Al-Dulaimy AI, Bunny MN, and Gray J (2010). Decision making under risk in swarm intelligence techniques. In Advances in Cognitive Systems. (Institution of Engineering and Technology), pp. 397–414.

3. Taddei F, Radman M, Maynard-Smith J, Toupance B, Gouyon PH, and Godelle B (1997). Role of mutator alleles in adaptive evolution. Nature 387, 700–702. [PubMed: 9192893]

4. Jetz W, and Rubenstein DR (2011). Environmental Uncertainty and the Global Biogeography of Cooperative Breeding in Birds. Current Biology 21, 72–78. [PubMed: 21185192]

5. Loewenstein GF, Weber EU, Hsee CK, and Welch N (2001). Risk as feelings. Psychological bulletin 127, 267. [PubMed: 11316014]

6. Jonzén N, Knudsen E, Holt RD, and Sæther B-E (2011). Uncertainty and predictability: the niches of migrants and nomads. Animal migration: a synthesis 91, 109.

7. Balázsi G, van Oudenaarden A, and Collins James J. (2011). Cellular Decision Making and Biological Noise: From Microbes to Mammals. Cell 144, 910–925. [PubMed: 21414483]

8. Jean-Pierre C, Marc P, Isabelle MB, and Bernard M (2000). Adaptation to the environment: Streptococcus pneumoniae, a paradigm for recombination-mediated genetic plasticity? Molecular Microbiology 35, 251–259. [PubMed: 10652087]

9. Claverys JP, Prudhomme M, Mortier-Barrière I, and Martin B (2000). Adaptation to the environment: Streptococcus pneumoniae, a paradigm for recombination-mediated genetic plasticity? Molecular microbiology 35, 251–259. [PubMed: 10652087]

10. Yu AJ, and Dayan P (2005). Uncertainty, neuromodulation, and attention. Neuron 46, 681–692. [PubMed: 15944135]

11. Platt ML, and Huettel SA (2008). Risky business: the neuroeconomics of decision making under uncertainty. Nature neuroscience 11, 398–403. [PubMed: 18368046]

12. Farashahi S, Donahue CH, Hayden BY, Lee D, and Soltani A (2019). Flexible combination of reward information across primates. Nature human behaviour 3, 1215–1224.

13. Blanchard TC, Hayden BY, and Bromberg-Martin ES (2015). Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. Neuron 85, 602–614. [PubMed: 25619657]

14. Gershman SJ, and Uchida N (2019). Believing in dopamine. Nature Reviews Neuroscience 20, 703–714. [PubMed: 31570826]

15. Soltani A, and Izquierdo A (2019). Adaptive learning under expected and unexpected uncertainty. Nature Reviews Neuroscience 20, 635–644. [PubMed: 31147631]

16. Gottlieb J (2012). Attention, learning, and the value of information. Neuron 76, 281–295. [PubMed: 23083732]

17. Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo CD, Tobler PN, and Bossaerts P (2008). Explicit neural signals reflecting reward uncertainty. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 363, 3801–3811. [PubMed: 18829433]

18. Hikosaka O, Yamamoto S, Yasuda M, and Kim HF (2013). Why skill matters. Trends in cognitive sciences 17, 434–441. [PubMed: 23911579]

19. Dolan RJ, and Dayan P (2013). Goals and habits in the brain. Neuron 80, 312–325. [PubMed: 24139036]

20. Wood W, and Rünger D (2016). Psychology of habit. Annual review of psychology 67, 289–314.

21. Shenhav A, Musslick S, Lieder F, Kool W, Griffiths TL, Cohen JD, and Botvinick MM (2017). Toward a rational and mechanistic account of mental effort. Annual review of neuroscience 40, 99–124.

22. Thorngate W (1976). Must we always think before we act? Personality and Social Psychology Bulletin 2, 31–35.

23. Graybiel AM (2008). Habits, rituals, and the evaluative brain. Annual review of neuroscience 31, 359–387.

24. Graybiel AM, and Smith KS (2014). Good habits, bad habits. Scientific American 310, 38–43.

25. Shiffrin RM, and Schneider W (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological review 84, 127.

26. Dayan P (2009). Goal-directed control and its antipodes. Neural networks : the official journal of the International Neural Network Society 22, 213–219. [PubMed: 19362448]

27. Courville AC, Daw ND, and Touretzky DS (2006). Bayesian theories of conditioning in a changing world. Trends in cognitive sciences 10, 294–300. [PubMed: 16793323]

28. Daw ND, O'doherty JP, Dayan P, Seymour B, and Dolan RJ (2006). Cortical substrates for exploratory decisions in humans. Nature 441, 876. [PubMed: 16778890]

29. ligaya K, Ahmadian Y, Sugrue LP, Corrado GS, Loewenstein Y, Newsome WT, and Fusi S (2019). Deviation from the matching law reflects an optimal strategy involving learning over multiple timescales. Nature communications 10, 1466.

30. Nassar MR, Wilson RC, Heasly B, and Gold JI (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. Journal of Neuroscience 30, 12366–12378. [PubMed: 20844132]

31. Doya K (2002). Metalearning and neuromodulation. Neural networks : the official journal of the International Neural Network Society 15, 495–506. [PubMed: 12371507]

32. Bromberg-Martin ES, and Hikosaka O (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. Neuron 63, 119–126. [PubMed: 19607797]

33. White JK, Bromberg-Martin ES, Heilbronner SR, Zhang K, Pai J, Haber SN, and Monosov IE (2019). A neural network for information seeking. Nature communications 10, 1–19.

34. Bennett D, Bode S, Brydevall M, Warren H, and Murawski C (2016). Intrinsic Valuation of Information in Decision Making under Uncertainty. PLoS computational biology 12, e1005020. [PubMed: 27416034]

35. Eliaz K, and Schotter A (2007). Experimental Testing of Intrinsic Preferences for NonInstrumental Information. American Economic Review 97, 166–169.

36. Kreps DM, and Porteus EL (1978). Temporal Resolution of Uncertainty and Dynamic Choice Theory. Econometrica 46, 185–200.

37. Wyckoff LB (1959). Toward a quantitative theory of secondary reinforcement. Psychological review 66, 68–78. [PubMed: 13645851]

38. Vasconcelos M, Monteiro T, and Kacelnik A (2015). Irrational choice and the value of information. Scientific reports 5, 13874. [PubMed: 26350951]

39. Charnov E, and Orians GH (2006). Optimal foraging: some theoretical explorations.

40. Adams GK, Watson KK, Pearson J, and Platt ML (2012). Neuroethology of decision-making. Current opinion in neurobiology 22, 982–989. [PubMed: 22902613]

41. Botvinick M, Ritter S, Wang JX, Kurth-Nelson Z, Blundell C, and Hassabis D (2019). Reinforcement learning, fast and slow. Trends in cognitive sciences.

42. Esber GR, and Haselgrove M (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: a model of attention in associative learning. Proceedings. Biological sciences / The Royal Society 278, 2553–2561.

43. Ma WJ, and Jazayeri M (2014). Neural coding of uncertainty and probability. Annual review of neuroscience 37, 205–220.

44. Zhang K, Chen CD, and Monosov IE (2019). Novelty, Salience, and Surprise Timing Are Signaled by Neurons in the Basal Forebrain. Current Biology 29, 134–142.e133. [PubMed: 30581022]

45. Monosov IE, Leopold DA, and Hikosaka O (2015). Neurons in the Primate Medial Basal Forebrain Signal Combined Information about Reward Uncertainty, Value, and Punishment Anticipation. The Journal of neuroscience : the official journal of the Society for Neuroscience 35, 7443–7459. [PubMed: 25972172]

46. Monosov IE, and Hikosaka O (2013). Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. Nature neuroscience 16, 756–762. [PubMed: 23666181]

47. White JK, and Monosov IE (2016). Neurons in the primate dorsal striatum signal the uncertainty of object-reward associations. Nature communications 7, 12735.

48. Monosov IE (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. Nature communications 8, 134.

49. O'Neill M, and Schultz W (2010). Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. Neuron 68, 789–800. [PubMed: 21092866]

50. Ledbetter MN, Chen DC, and Monosov IE (2016). Multiple mechanisms for processing reward uncertainty in the primate basal forebrain. The Journal of neuroscience : the official journal of the Society for Neuroscience 36.

51. Ghazizadeh A, Griggs W, and Hikosaka O (2016). Object-finding skill created by repeated reward experience. Journal of vision 16, 17.

52. Farashahi S, Donahue CH, Khorsand P, Seo H, Lee D, and Soltani A (2017). Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. Neuron 94, 401–414.e406. [PubMed: 28426971]

53. Khorsand P, and Soltani A (2017). Optimal structure of metaplasticity for adaptive learning. PLoS computational biology 13, e1005630. [PubMed: 28658247]

54. Horan M, Daddaoua N, and Gottlieb J (2019). Parietal neurons encode information sampling based on decision uncertainty. Nature neuroscience, 1.

55. Rouhani N, Norman KA, and Niv Y (2018). Dissociable effects of surprising rewards on learning and memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 44, 1430.

56. Paton JJ, and Buonomano DV (2018). The Neural Basis of Timing: Distributed Mechanisms for Diverse Functions. Neuron 98, 687–705. [PubMed: 29772201]

57. Wang J, Narain D, Hosseini EA, and Jazayeri M (2018). Flexible timing by temporal scaling of cortical responses. Nature neuroscience 21, 102. [PubMed: 29203897]

58. MacDonald CJ, Lepage KQ, Eden UT, and Eichenbaum H (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. Neuron 71, 737–749. [PubMed: 21867888]

59. Raver SM, and Lin S-C (2015). Basal forebrain motivational salience signal enhances cortical processing and decision speed. Frontiers in behavioral neuroscience 9.

60. Mesulam MM, Mufson EJ, Levey AI, and Wainer BH (1983). Cholinergic innervation of cortex by the basal forebrain: cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (substantia innominata), and hypothalamus in the rhesus monkey. The Journal of comparative neurology 214, 170–197. [PubMed: 6841683]

61. Turchi J, Saunders RC, and Mishkin M (2005). Effects of cholinergic deafferentation of the rhinal cortex on visual recognition memory in monkeys. Proceedings of the National Academy of Sciences of the United States of America 102, 2158–2161. [PubMed: 15684066]

62. Griggs WS, Kim HF, Ghazizadeh A, Gabriela Costello M, Wall KM, and Hikosaka O (2017). Flexible and Stable Value Coding Areas in Caudate Head and Tail Receive Anatomically Distinct Cortical and Subcortical Inputs. Frontiers in neuroanatomy 11, 106. [PubMed: 29225570]

63. Hangya B, Ranade SP, Lorenc M, and Kepecs A (2015). Central Cholinergic Neurons Are Rapidly Recruited by Reinforcement Feedback. Cell 162, 1155–1168. [PubMed: 26317475]

64. Lin SC, and Nicolelis MA (2008). Neuronal ensemble bursting in the basal forebrain encodes salience irrespective of valence. Neuron 59, 138–149. [PubMed: 18614035]

65. Turchi J, Chang C, Frank QY, Russ BE, David KY, Cortes CR, Monosov IE, Duyn JH, and Leopold DA (2018). The basal forebrain regulates global resting-state fMRI fluctuations. Neuron 97, 940–952. e944. [PubMed: 29398365]

66. Reimer J, McGinley MJ, Liu Y, Rodenkirch C, Wang Q, McCormick DA, and Tolias AS (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. Nature communications 7, 1–7.

67. Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, and Gold JI (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. Nature neuroscience 15, 1040. [PubMed: 22660479]

68. Dasilva M, Brandt C, Gotthardt S, Gieselmann MA, Distler C, and Thiele A (2019). Cell class-specific modulation of attentional signals by acetylcholine in macaque frontal eye field. Proceedings of the National Academy of Sciences 116, 20180–20189.

69. Thiele A, and Bellgrove MA (2018). Neuromodulation of attention. Neuron 97, 769–785. [PubMed: 29470969]

70. Sajedin A, Menhaj MB, Vahabie A-H, Panzeri S, and Esteky H (2019). Cholinergic modulation promotes attentional modulation in primary visual cortex-a modeling study. Scientific reports 9, 1–18. [PubMed: 30626917]

71. Monosov IE, and Thompson KG (2009). Frontal eye field activity enhances object identification during covert visual search. Journal of neurophysiology 102, 3656–3672. [PubMed: 19828723]

72. Noudoost B, Chang MH, Steinmetz NA, and Moore T (2010). Top-down control of visual attention. Current opinion in neurobiology 20, 183–190. [PubMed: 20303256]

73. Schafer RJ, and Moore T (2007). Attention governs action in the primate frontal eye field. Neuron 56, 541–551. [PubMed: 17988636]

74. Ghazizadeh A, Griggs W, and Hikosaka O (2016). Ecological origins of object salience: Reward, uncertainty, aversiveness, and novelty. Frontiers in neuroscience 10, 378. [PubMed: 27594825]

75. Daddaoua N, Lopes M, and Gottlieb J (2016). Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. Scientific reports 6, 20202. [PubMed: 26838344]

76. Krauzlis RJ, Lovejoy LP, and Zénon A (2013). Superior Colliculus and Visual Spatial Attention. Annual review of neuroscience 36, 10.1146/annurev-neuro-062012-170249.

77. Yanike M, and Ferrera VP (2014). Representation of outcome risk and action in the anterior caudate nucleus. The Journal of neuroscience : the official journal of the Society for Neuroscience 34, 3279–3290. [PubMed: 24573287]

78. Hikosaka O (2007). Basal ganglia mechanisms of reward-oriented eye movement. Annals of the New York Academy of Sciences 1104, 229–249. [PubMed: 17360800]

79. Hong S, and Hikosaka O (2008). The globus pallidus sends reward-related signals to the lateral habenula. Neuron 60, 720–729. [PubMed: 19038227]

80. Hong S, Jhou TC, Smith M, Saleem KS, and Hikosaka O (2011). Negative reward signals from the lateral habenula to dopamine neurons are mediated by rostromedial tegmental nucleus in primates. Journal of Neuroscience 31, 11457–11471. [PubMed: 21832176]

81. Monosov IE, and Hikosaka O (2012). Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience 32, 10318–10330. [PubMed: 22836265]

82. Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, and Botvinick M (2020). A distributional code for value in dopamine-based reinforcement learning. Nature, 1–5.

83. Behrens TE, Woolrich MW, Walton ME, and Rushworth MF (2007). Learning the value of information in an uncertain world. Nature neuroscience 10, 1214–1221. [PubMed: 17676057]

84. Pushkarskaya H, Liu X, Smithson M, and Joseph JE (2010). Beyond risk and ambiguity: Deciding under ignorance. Cognitive, Affective, & Behavioral Neuroscience 10, 382–391.

85. Li YS, Nassar MR, Kable JW, and Gold JI (2019). Individual neurons in the cingulate cortex encode action monitoring, not selection, during adaptive decision-making. Journal of Neuroscience 39, 6668–6683. [PubMed: 31217329]

86. Sarafyazd M, and Jazayeri M (2019). Hierarchical reasoning by neural circuits in the frontal cortex. Science 364, eaav8911. [PubMed: 31097640]

87. Seymour B, Singer T, and Dolan R (2007). The neurobiology of punishment. Nature reviews. Neuroscience 8, 300–311. [PubMed: 17375042]

88. Worbe Y, Palminteri S, Savulich G, Daw ND, Fernandez-Egea E, Robbins T, and Voon V (2016). Valence-dependent influence of serotonin depletion on model-based choice strategy. Molecular psychiatry 21, 624–629. [PubMed: 25869808]

89. Kojima S, Kao MH, Doupe AJ, and Brainard MS (2018). The avian basal ganglia are a source of rapid behavioral variation that enables vocal motor exploration. Journal of Neuroscience 38, 9635–9647. [PubMed: 30249800]

90. Neuringer A (2002). Operant variability: Evidence, functions, and theory. Psychonomic Bulletin & Review 9, 672–705. [PubMed: 12613672]

91. Trageser JC, Monosov IE, Zhou Y, and Thompson KG (2008). A perceptual representation in the frontal eye field during covert visual search that is more reliable than the behavioral report. The European journal of neuroscience 28, 2542–2549. [PubMed: 19032593]

92. Messinger A, Squire LR, Zola SM, and Albright TD (2005). Neural correlates of knowledge: stable representation of stimulus associations across variations in behavioral performance. Neuron 48, 359–371. [PubMed: 16242414]

93. Wittmann MK, Kolling N, Akaishi R, Chau BK, Brown JW, Nelissen N, and Rushworth MF (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nature communications 7, 12327.

94. Kim HF, and Hikosaka O (2013). Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. Neuron 79, 1001–1010. [PubMed: 23954031]

95. Kim HF, Ghazizadeh A, and Hikosaka O (2014). Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. Frontiers in neuroanatomy 8, 120. [PubMed: 25400553]

96. Whittington JC, and Bogacz R (2019). Theories of error back-propagation in the brain. Trends in cognitive sciences.

97. Gottlieb J, Oudeyer P-Y, Lopes M, and Baranes A (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends in cognitive sciences 17, 585–593. [PubMed: 24126129]

98. Haber N, Mrowca D, Wang S, Fei-Fei LF, and Yamins DL (2018). Learning to play with intrinsically-motivated, self-aware agents. In Advances in Neural Information Processing Systems. pp. 8388–8399.

99. Zambaldi V, Raposo D, Santoro A, Bapst V, Li Y, Babuschkin I, Tuyls K, Reichert D, Lillicrap T, and Lockhart E (2018). Relational deep reinforcement learning. arXiv preprint arXiv:1806.01830.

BOX 1 –

Uncertainty about aversive events

This review has concerned itself with uncertainty about rewards. However, recent evidence from our laboratory indicates that there may be distinct circuitry that processes uncertainty about the occurrence of aversive or noxious events, that we refer to as punishment uncertainty [48]. Despite the fact that punishment uncertainty is a crucial variable in the control of emotional states such as anxiety and depression, at the moment, little is known about the neural circuits of punishment uncertainty and about their impact on behavior.

Uncertainty about appetitive outcomes and about aversive punishments is associated with gains and losses. For example, following reward uncertainty, a gain could be a reward, and following punishment uncertainty, a gain could be an omission of a noxious event. However, these positive reward prediction errors are associated, not only with distinct outcomes, but also with distinct emotional states: relief when punishment is omitted and joy when reward is delivered [87]. Diverse responses while anticipating and receiving rewarding and aversive outcomes elicit different learning strategies [88], exert differential influences on choice related computations [5], and promote distinct action plans (e.g., fight or flight versus stay or leave) [87]. These observations suggest that the brain may contain distinct circuits for processing reward and punishment uncertainty.

Our laboratory recently discovered that the ACC contains distinct populations of reward and punishment uncertainty sensitive neurons [48]. Particularly, a small sub region of anterior ventral ACC, above the corpus callosum, is enriched in single neurons that respond to uncertain predictions of punishments, but not to uncertain predictions of rewards [48].

Interestingly, our experiments failed to identify punishment uncertainty signals in many brain regions that are involved in reward uncertainty related computations [45-47]). Hence, it is likely that the ACC sends punishment- and reward- uncertainty signals to distinct brain regions to support divergent processes, such as for example to separately control the desire to seek advance information about pleasant versus aversive outcomes, and to differentially impact choice behavior [5]. It will therefore be crucial to identify the brain regions to which the ACC sends punishment uncertainty signals, and assess how they utilize punishment uncertainty signals to mediate the distinct effects of reward and punishment on cognition and mood [5].

BOX 2 –

Learning by artificial learners

Here I review several observations from the studies of neurobiology and behavior under uncertainty that may be relevant in the context of designing artificial intelligence systems.

Beyond the fact that biological agents explicitly represent uncertainty, they are in fact intrinsically motivated to spend time and effort exploring their environment. In many regimes, their cost functions bias them towards information acquisition and uncertainty reduction through surprise and novelty seeking. Particularly, mammals prior to sexual maturity, when their learning rates are high and their brains exceptionally plastic, tend to explore their environment extensively, often under the guidance of adults. This relatively long training-period in which novelty and surprise have a profound effect on behavior may give them an advantage later in life or in situations where accurate predictions of action values are crucial for survival.

Exploration or "play" based learning may be particularly enhanced in mammals that display sophisticated social structures and hierarchies, including primates. Living in groups imposes not only additional sources of uncertainty and opportunities for learning (because you don't know what the other group members will do) but also the need to cooperate. As in the case of children at the playground under the watchful eyes of adults, multi-agent cooperation can support curiosity / information seeking behaviors in younger agents as they explore the world within the constraints of others' more value-based policies.

Another way in which exploration and behavioral variability can arise in biological agents is through noise or variability in neuronal computations across a wide range of circuits. The basal ganglia for example produces variability in action selection [89] to enhance exploration and learning of new or arbitrary associations [90]. Since animals have distributed systems for sensory, motor, and cognitive functions, exploration could result from noise and/or errors at many levels of information processing [91, 92] in a context-dependent manner.

Also, an important feature of biological systems to consider for artificial learning under uncertainty is that the brain contains many neuronal circuits that have distinct learning rates and learn from distinct teaching signals [93-95]. These circuits changed through evolution (or development), sometimes together, sometimes at different time scales to cooperate or oppose each other. This heterogeneity can allow animals to learn flexibly, in a context-dependent manner (and itself can be a function of a higher-level objective). For example, a child may explore language fluidly to maximize surprise and learning, while still maintaining strong and stable aversion to certain foods she tried only once. Multiple teaching signals such as prediction errors signaled by global neuromodulators (or errors computed more locally during sensory processing [96]) can support such flexibility and diversity in learning, on different time scales [95].

There is appreciable work to implement the motivation to explore and to seek information in artificial neural networks [97-99]. Thus far, most of the efforts are based on designing

cost functions that heavily weight novelty and surprise to help agents learn the structure of the world [98]. Designing networks that represent multiple agents having different goals may be important for this effort. For example, in adversarial learning one agent is trained by challenges from another agent whose goal is to "fool" the first agent and maximize its error rate. Another way for implementing multi-agent processes in artificial neural networks could be for an agent to maximize others' novelty and surprise, rather than error. Moreover, just as artificial-learners can be designed to be curious about the reward-seeking strategies of other agents and learn by observing their examples, it may be beneficial to also learn by observing others' information-seeking strategies.

If artificial-agents are to gain qualities of "intelligent" behavior they will likely need to comprise many networks learning at different rates, for distinct goals. And, to coordinate them, one future avenue may be to add controllers of global *network states.* This would allow neural networks to follow the examples of biological agents that intelligently switch between different modes of network coordination and function based on what is most efficient for their current situation. For instance, responding to a lull in their opportunities to gain reward, agents may sustain adaptive learning by temporarily suppressing action, and enhancing memory replay and other deliberative functions to continue learning from past experiences.

Outstanding questions

How do agents use their internal models about how volatile the environment is and about how much of the variability in outcomes is due to inherent variance or measurement noise to mediate learning?

Are there distinct circuits that selectively signal reward uncertainty under risk versus ambiguity? If not, how might the brain use representations of reward uncertainty differentially in different contexts?

Does uncertainty influence episodic memory distinctly from prediction errors?

Uncertainty is followed by surprise. Both reward uncertainty and surprise influence learning, particularly the learning of action values. But to date, little is known about how uncertainty and surprise influence other types of learning, such as of the encoding and retrieval of episodic memory.

What is the relationship among risk attitude, information seeking, and willingness to endure mental effort? What are the neural circuits that mediate their interactions during value-based decision making? What quantitative measures of uncertainty affect risk attitude and information seeking?

How does uncertainty promote or suppress automatic and controlled behaviors?

What is the cellular identity of phasic and ramping basal forebrain neurons and what is their connectivity with, and effect on, different cortical computational subunits?

The anterior cingulate (ACC) – basal ganglia network has been linked to driving information seeking to resolve uncertainty. What is the source of uncertainty signals in this network? The ACC displays correlations with information seeking-related gaze behavior earlier than the basal ganglia. But, within this network, uncertainty signals first emerge as a rapid suppression of pallidal neurons deep within the basal ganglia circuitry. Given this observation, future research should identify the sources of uncertainty signals in the basal ganglia.

Are information seeking strategies and mechanisms effector specific, for instance among arm reaches versus gaze shifts? Given that in cortex and in the basal ganglia, hand/reach control regions are somewhat separated from oculomotor control regions, inputs from the ACC-basal ganglia network could impact effectors differentially, which may be helpful to flexibly control behavior depending on goal, context, the physical parameters, and costs of individual actions.
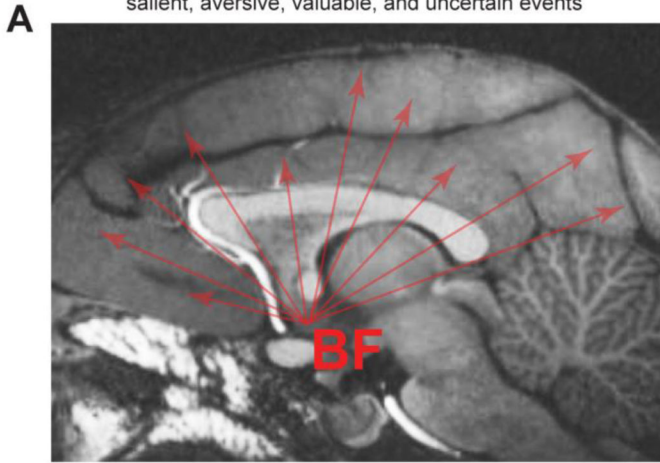
Highlights

- Uncertainty about future outcomes mediates attention, learning, memory and decision making. Uncertainty can modulate, for example, choices on whether to make risky decisions or to prioritize actions aimed at gaining uncertainty-reducing information

- The basal forebrain broadcasts information about uncertainty and surprise to guide learning, memory, and attention

- A cotrico-basal ganglia loop originating in the anterior cingulate guides information seeking about uncertain rewards

- Ongoing work is assessing how neural circuits generate, support, and implement the mental algorithms that govern uncertainty-related behaviors

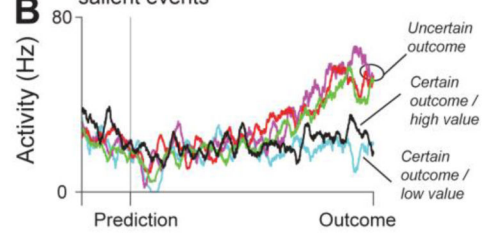|	Uncertainty low	Uncertainty high
Behavioral policy	Exploit	Explore
Goal of learning	Adjust behavior to increase reward per unit time	Obtain model of environment
Value of uncertainty resolving information	Low	High
Episodic memory	low	high
Cognition and attention	low	high
Controlled (vs. automatic) behavior	low	high

Figure 1. The goals of behavior change when uncertainty is high.

Tracking reward uncertainty is crucial for adaptive control of a wide range of behavioral and learning strategies (top two rows). Under high reward uncertainty it is important to obtain a model of the environment and reduce uncertainty, hence serving to increase confidence in action values and in upcoming outcomes. Under low reward uncertainty, if reward is available, fast automatic behaviors can be deployed and dynamically optimized through model-free learning to increase reward per unit time. The process of uncertainty reduction can be supported by assigning a higher value to uncertainty-reducing information when uncertainty is high, and/or mediating risk attitude (not shown, but discussed in Main Text). Learning and information seeking under uncertainty are supported by increases in demand on memory and attention, and are associated with increases in the need for online evaluation of and *control* over many actions and behaviors. Of note, though, this does not imply that in high reward uncertainty, automatic behaviors and related learning processes are suppressed. In fact, high capacity for automating some behaviors may support primates' high capacity to explore and obtain multi-dimensional models of the world through mental processes such as deliberation (see Concluding Remarks).
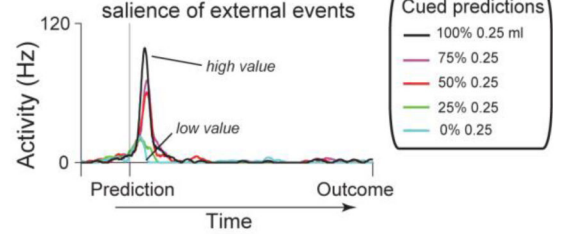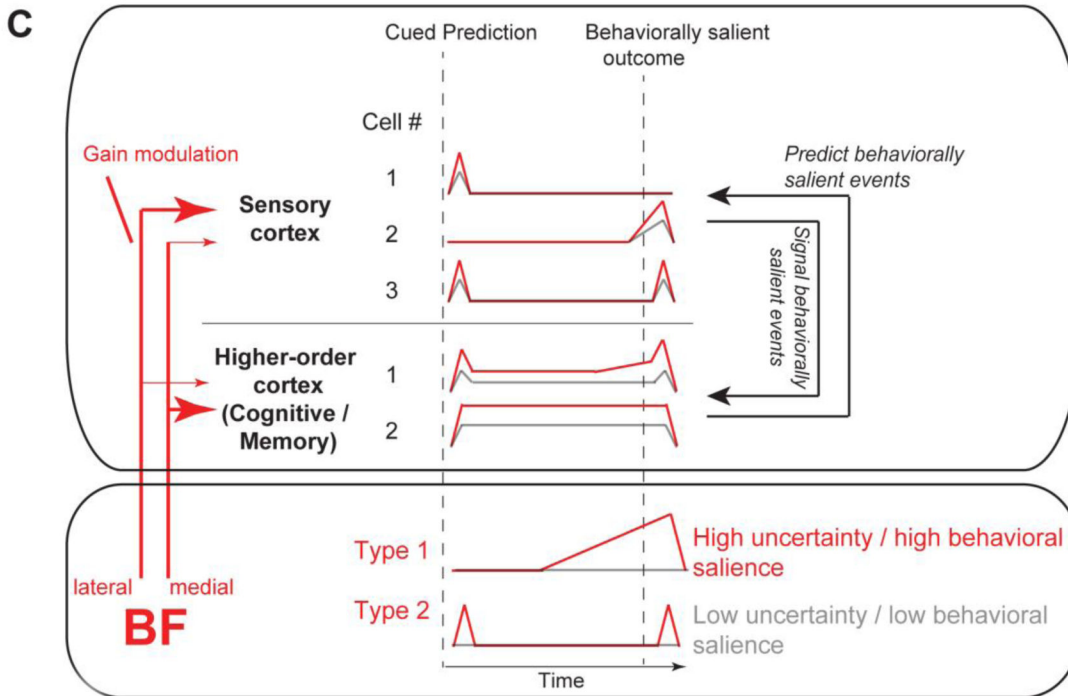
Figure 2. Temporally-precise uncertainty-related modulation of wide-scale cortical computations by the basal forebrain (BF).

(**A-C**) BF broadcasts the timing and predictions of behaviorally salient, aversive, valuable, and uncertain events [44, 45, 63, 64]. (**A**) Location of the basal forebrain shown on sagittal MRI of a macaque monkey brain. Red arrows represent the wide projections of BF neurons to the cortical mantle. MR image is reproduced from reference [45]. (**B**) Two functional classes of BF neurons: "Type 1" neurons (top), which predict uncertain salient events, and "Type 2" neurons (bottom) that encode these events' salience and timing, after they occur. The neurons' average activity is separated to trials in which five distinct probabilistic reward

predictions were made with five distinct visual cues. The outcomes (reward or no reward) occurred 2.5 seconds after each prediction. Both ramping (Type 1) and phasic (Type 2) neurons are found in the medial and lateral regions of the BF, though the medial BF is particularly enriched with ramping (Type 1, top) neurons (reproduced from [45]). (**C**) Temporally precise modulation and coordination of cortical activity by the BF, constrained by anatomical biases in BF connectivity to distinct functional regions of the cortical mantle [60, 65]. Thickness of red arrows from the medial and lateral BF denotes connectivity strength (red arrows, on left). Cartoon of example cells' responses in sensory and higher-order cortices (top square) with (red) and without (gray) BF-mediated gain modulation to external events that predict behaviorally/motivationally salient outcomes. This temporally precise gain enhancement across distinct cortical subregions may be crucial for rapid control of cognitive functions under uncertainty (Figure 1), such as prediction-mediated control of sensory processing resembling attention (black arrows on the right in top square). Furthermore, wide-scale temporally precise gain modulation may mediate spike-timing across many cortical areas to impact slower time scale changes for learning and memory (e.g. by enhancing prediction error or external salient event responses by ramping activity of Type 1 neurons).
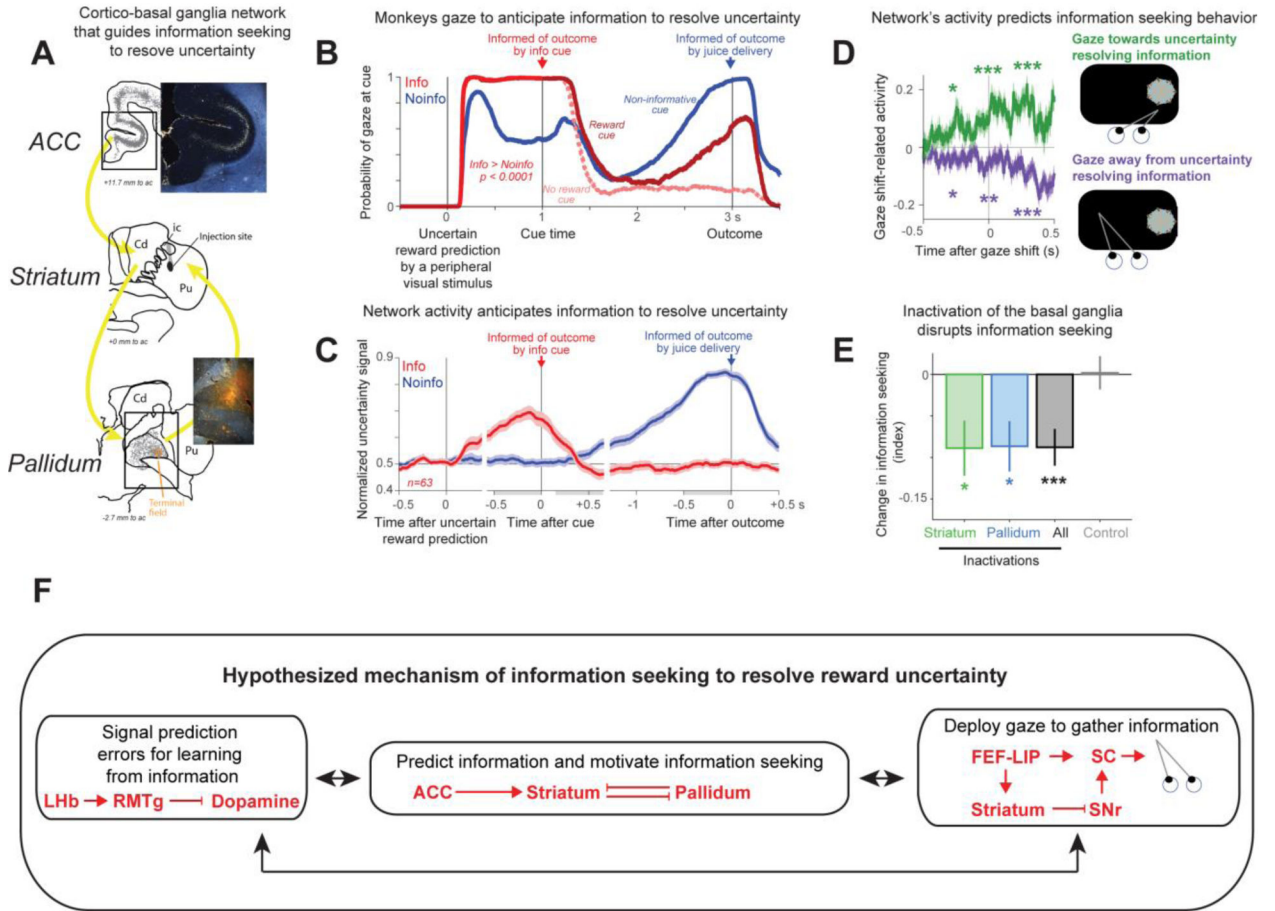
Figure 3. A primate cortical basal ganglia network for information seeking.
(**A**) Example bidirectional tracer injection into an uncertainty-related region of the striatum reveals connectivity with anterior cingulate and anterior ventral pallidum. (**B**) Monkeys' uninstructed gaze behavior in an information observation task reveals motivation to seek advance information about uncertain rewards. On Info trials (red), a peripheral visual stimulus predicted uncertain rewards. One second after, it was replaced by uncertainty-resolving cue stimuli (see red arrow). Monkeys' gaze on Info trials was attracted to the location of the uncertain prediction in anticipation of receiving informative cues that resolved their uncertainty. After uncertainty resolution in Info trials, gaze is split to trials in which reward will be delivered (dark red) and reward was not delivered (pink). On Noinfo trials (blue), another visual cue also predicted uncertain rewards. Here, the subsequent cue stimuli, shown 1 second after, were not informative and the monkeys resolved their uncertainty at the time of the trial outcome (blue arrow). In Noinfo trials, gaze was particularly attracted to uncertain visual stimuli in anticipation of uncertainty resolution by outcome delivery. Probability of gazing at the stimulus ramped up in anticipation of the uncertain outcome until it became greater than other conditions (50% > 100%; compare blue with dark red). (**C**) Neural activity across the ACC-Striatum-Pallidum network anticipates uncertainty resolution. Same format as (B). (**D**) The networks' activity predicts information seeking gaze behavior. Mean time course of gaze shift-related normalized network activity

aligned on gaze shifts onto the uncertainty resolving informative stimulus (green) and off the stimulus (purple). Asterisks indicate significance in time windows before, during, and after the gaze shift. Uncertainty-related activity was significantly enhanced before gaze shifts on the stimulus and reduced before gaze shifts off the stimulus when animals were anticipating information about an uncertain reward. (E) Pharmacological inactivation of the striatum and pallidum regions enriched with information anticipating neurons disrupted information seeking relative to saline control which produced no effect. (**F**) Mechanisms (black) and circuits (red) that support information seeking. BF is not shown but it projects to all cortical sites shown in F. ACC – anterior cingulate cortex; Cd – Caudate; FEF - Frontal eye field, LHb – lateral habenula complex; LIP – lateral intraparietal cortex, Pu – Putamen; RMTg - rostromedial tegmental nucleus; SC – superior colliculus; SNr – substantia nigra pars reticulata; ac – anterior commissure. The images in A-E are reproduced from [33].