Neuroscience Research 178 (2022) 10–19

Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.sciencedirect.com/journal/neuroscience-research

Review Article

Metacognitive resources for adaptive learning★ [☆]

Aurelio Cortese

Computational Neuroscience Labs, ATR Institute International, 619-0288 Kyoto, Japan

ARTICLE INFO

Keywords:
Confidence
Metacognition
Abstraction
Learning
Reinforcement learning
Meta-representations
Multivoxel neural reinforcement
Decoded neurofeedback

ABSTRACT

Biological organisms display remarkably flexible behaviours. This is an area of active investigation, in particular in the fields of artificial intelligence, computational and cognitive neuroscience. While inductive biases and broader cognitive functions are undoubtedly important, the ability to monitor and evaluate one's performance or oneself – *meta*cognition – strikes as a powerful resource for efficient learning. Often measured as decision confidence in neuroscience and psychology experiments, metacognition appears to reflect a broad range of abstraction levels and downstream behavioural effects. Within this context, the formal investigation of how metacognition interacts with learning processes is a recent endeavour. Of special interest are the neural and computational underpinnings of confidence and reinforcement learning modules. This review discusses a general hierarchy of confidence functions and their neuro-computational relevance for adaptive behaviours. It then introduces novel ways to study the formation and use of meta-representations and nonconscious mental representations related to learning and confidence, and concludes with a discussion on outstanding questions and wider perspectives.

The ability to learn *efficiently* is a fascinating aspect of biological intelligence. Humans can learn new tasks or behaviours in an often predictably fast manner – with limited data. Machines can reach and periodically surpass human performance in specific contexts [e.g., Atari games, or chess (Lake et al., 2015; Sorokin et al., 2015)]. Yet, the way in which humans and machines learn and approach problems remains (for the most part) strikingly different, and particularly so for complex behaviours. Recent empirical and theoretical work suggests inductive abilities (Lake et al., 2015; Tenenbaum et al., 2011) and higher cognitive functions such as attention, episodic memory as well as metacognition and consciousness, may be crucial ingredients to augment artificial agents and accelerate learning (Behrens et al., 2018; Bengio, 2017; Botvinick et al., 2019; Cortese et al., 2019; Niv, 2019).

The focus of this review will be on the intersection between metacognition and learning. Metacognition, or cognition about cognition, is the ability to reflect upon and report one's own mental states (Fleming et al., 2012; Metcalfe and Son, 2012). Interestingly, although many other animals have some form of metacognition, it appears to be uniquely developed in humans (Metcalfe, 2008). Given its self-monitoring nature, metacognition is well positioned to integrate feedback loops over behaviour and influence learning processes. It could take part in generating *meta*-representations – summarised *re*-representations of ongoing sensory or memory representations (Brown et al., 2019; Dehaene et al., 2017; Lau and Rosenthal, 2011). This feature seems inherently important in a functional architecture of hierarchical learning systems such as the brain (Kawato et al., 1987; Wang et al., 2018), since it provides direct substrate for computations at high abstraction levels. Recent work has begun to unravel how metacognition interacts with reinforcement learning, and how this interaction could be one key solution to learn flexible behavioural policies in complex and noisy environments (Cortese et al., 2021, 2020; Lak et al., 2020; Lebreton et al., 2019). In this paper I will briefly cover how metacognition is operationalized, the neural and computational underpinnings, to then focus on its relevance for learning algorithms. Finally, I will introduce novel ways to study neural representations and meta-representations related to learning and metacognition, and discuss outstanding questions and directions for future research.

1. Metacognition: from perception and memory to high level reasoning and strategies

How do we operationalize metacognition? In neuroscience and psychology, metacognition is often measured through confidence. Yet 'confidence' is remarkably broad. It can relate to one's overall self-confidence (I am confident that I will do well in today's undertakings), to decisions (I don't know whether choosing to turn left was

[☆] Review article for the Japan Neuroscience Society Young Investigator Award 2021.
E-mail address: cortese_a@atr.jp.

https://doi.org/10.1016/j.neures.2021.09.003
Received 30 March 2021; Received in revised form 7 September 2021; Accepted 8 September 2021
Available online 15 September 2021

the correct choice), perception (it is foggy, and I am unsure whether what I see in the distance is an animal or an object), or memory (I am certain I locked the door when I left).

In human experiments, confidence is generally recorded as an explicit judgement / report, through likert rating scales (Rahnev et al., 2020). More sophisticated metrics may be used, such as wagering on the outcome of a choice (Kepecs and Mainen, 2012; Persaud et al., 2007). In other animals, in the absence of verbal reports, implicit measures are more common. Studies in monkeys tend to use opt-out tasks, whereby the animal can choose a safe option instead of a more rewarding but riskier discrimination (Kiani and Shadlen, 2009; Komura et al., 2013). The intuition behind this approach is that with low confidence, the monkey will choose to opt-out more often. In rodents, similar wagering paradigms are used, where confidence is indexed as the animal's willingness to wait for a reward (Kepecs et al., 2008; Miyazaki et al., 2018; Stolyarova et al., 2019).

It is worthwhile to make a distinction at this point on the definition of confidence, as research has partially evolved along parallel trajectories. Work in animals has privileged a probabilistic view of confidence, with strong computational undertones. In this normative definition, confidence directly reflects the sensory evidence, or the noisy probability that the decision was correct (Kepecs et al., 2008; Kiani and Shadlen, 2009; Meyniel et al., 2015). Human studies have, instead, tended to incorporate broader views, focusing on both normative definitions of confidence (Sanders et al., 2016) as well as on psychological, cognitive aspects grounded in signal detection theory (Fleming and Lau, 2014; Rounis et al., 2010). Nevertheless, the past few years have seen a blur of this cross-species distinction, with substantial benefits for the field. In terms of the terminology used – e.g., certainty vs confidence (Dehaene et al., 2017; Pouget et al., 2016), the computational approaches used to determine the basis of confidence judgements (Maniscalco et al., 2021; Meyniel et al., 2015; Peters and Lau, 2015; Pouget et al., 2016), and on the crucial yet overlooked issue of confounding variables (Fleming and Lau, 2014; Maniscalco and Lau, 2016; Morales et al., 2019).

The eclectic range of confidence definitions and measures leaves us with many unanswered questions, despite decades of research. Is reported confidence a simple translation of statistical confidence? Similarly, is confidence unidimensional or are there multiple signals converging into a final experience and reported judgement? At the neural level, is there a core brain circuit responsible for primary self-monitoring functions? Or rather, are there basic neuro-computational motifs repeated throughout the brain, each module tuned to different aspects of monitoring / evaluation? Answering these questions will help us understand the function of confidence in learning. Behavioural and neuroimaging studies in humans, as well as electrophysiological studies in non-human primates and rodents, have begun to resolve these questions. Research has identified neural correlates of confidence operating in a single domain (e.g., perceptual only vs. memory only), as well as more abstract (e.g., both perception and memory) (McCurdy et al., 2013; Miyamoto et al., 2017; Morales et al., 2018). While there are distinct neuro-anatomical substrates of confidence at different levels of abstraction, the underlying computations may not differ drastically – although this remains to be shown.

From a computational standpoint, several recent studies have found confidence to interact with reinforcement learning (Cortese et al., 2020; Lak et al., 2019, 2017), as well as guiding future choices (De Martino et al., 2012; Folke et al., 2016), and operating on reasoning / credit assignment (Sarafyazd and Jazayeri, 2019). Intriguing perspectives are arising, and a closer look at the computational value of confidence will help lay the ground for future investigations on the very nature of adaptive behaviours.

2. The computational value of metacognition and confidence in reinforcement learning

Intuitively, confidence reflects the certainty or uncertainty in one's decision, skill, or knowledge. Keeping internal certainty / uncertainty signals can be useful to shape one's future choices (Folke et al., 2016) – e.g., if I was not confident and the outcome was negative, next time I might consider another option; or in controlling how much an agent needs to learn (the learning rate) (Nassar et al., 2010) – e.g., I am confident, therefore I don't need drastic updates in my beliefs. These examples provide a first layer of qualitative evidence for the adaptive function of confidence.

To establish how confidence can affect learning (and vice-versa too, how e.g., choice outcomes may affect confidence) in a quantitative manner, we need the formalism of mathematical models. In this context, reinforcement learning is a very successful framework that explains learning and particularly, learning from experience (Doya, 2007; Sutton and Barto, 1998). In classic reinforcement learning, agents learn a policy through reward and / or punishment. Essentially, learning a set of conditions (states), under which certain actions are better than others (e.g., more likely to lead to reward). While this is a drastic simplification of the interactions any biological organism will have with their environment, reinforcement learning encapsulates a set of seemingly universal rules (Dayan and Niv, 2008; Sutton and Barto, 1998). Formally, reinforcement learning is described by a set of 'states' (the description of the environment, such as conditions, locations, etc), 'actions' (what the agent can do, such as choosing between options A, B, C, or moving right, left, front, back, etc), and 'outcomes' (whether the action taken results in a small or big reward, punishment, etc). The agent therefore learns – often through a sort of look-out table called the *Q* value function (Watkins and Dayan, 1992) – a policy that maximizes long-term returns by taking the most valuable action given a certain state. It is important to stress here this notion of policy; the best action might not necessarily be the one that leads to the highest immediate return, but the one that would do so in the long-term. While there exist a panoply of more complex models, two basic parameters govern the way an agent learns through reinforcement. The learning rate (α) controls how much the mismatch between an outcome and the agent's prediction (the prediction error [PE]) will affect the agent's value estimates. Stochasticity (β, inverse temperature) instead regulates the amount of 'randomness' in the agent's action selection (how much the value estimates will influence action selection). We will next look at how confidence intersects with learning (Fig. 1A), from the accumulation of evidence to model parameters, value estimates, prediction errors and model selection.

2.1. Evidence accumulation

A crucial step in any learning scenario is to acquire the right amount of information. Confidence has been shown to play a key function in controlling evidence accumulation online (Balsdon et al., 2020; Brosnan et al., 2020; Lim et al., 2020). Intuitively, if I am already sure that my decision, or action, is correct, then there is no need for me to spend additional time / resources accumulating more evidence. It is worth mentioning that evidence accumulation can mean momentary evidence (subsecond range, accumulating evidence for a single choice), as well as evidence over longer time scales across entire episodes.

2.2. Learning rate

Confidence can control the learning rate - low confidence would engage higher learning rates because little is known about the environment or the current situation (Nassar et al., 2012, 2010). High confidence would have the opposite effect whereby the learning rate will shrink - if an agent is confident about its choices, behaviour, or strategy, then there should be little need to cause large updates to one's value estimates. More recent theoretical work has further shown how the learning rate is proportional (inversely proportional) to confidence for incorrect (correct) decisions (Drugowitsch et al., 2019). Importantly, an open question is whether these effects take place at global and / or local levels, depending on the learning horizon and environment factors (such

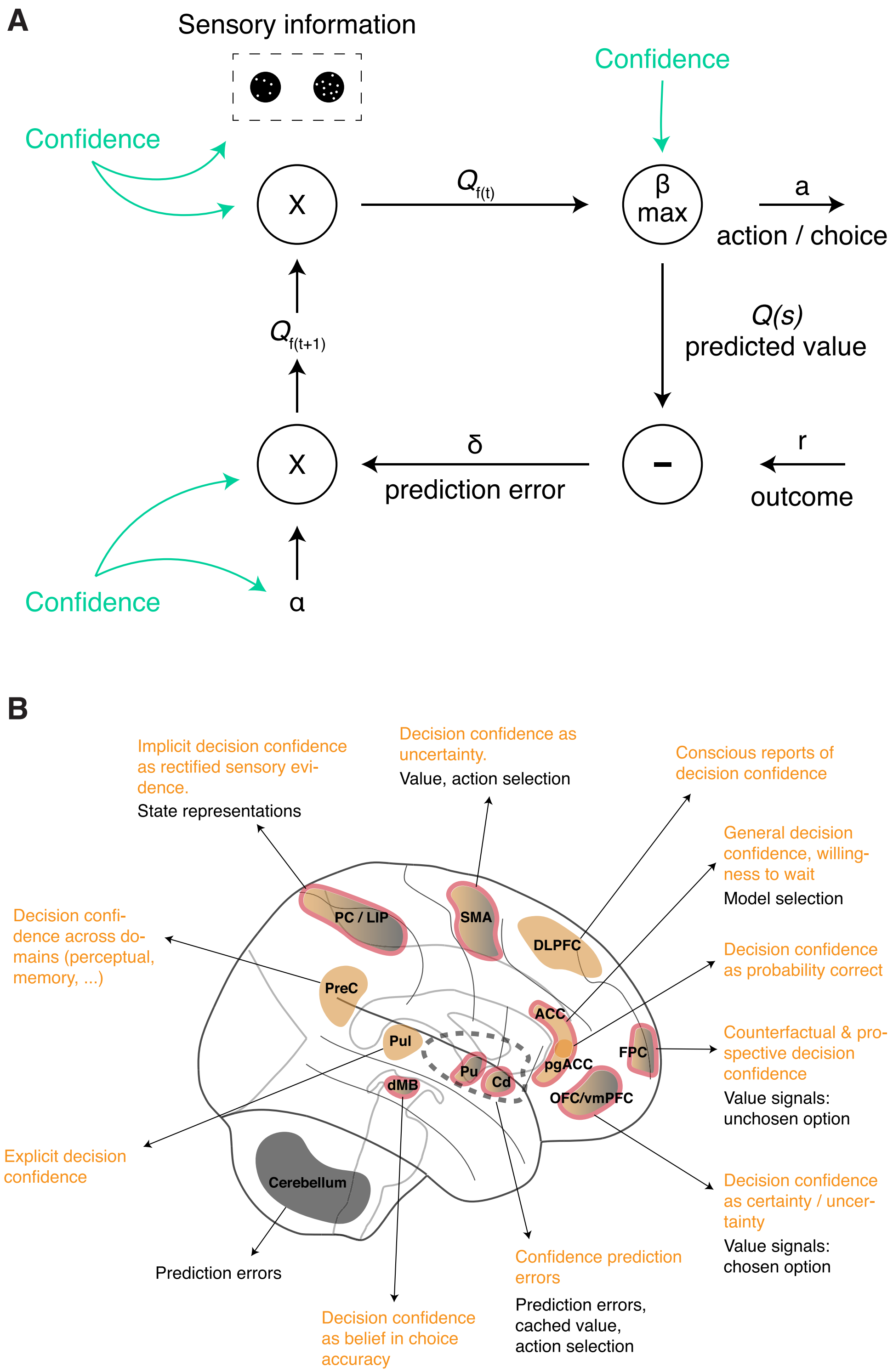
A
Sensory information
Confidence
Confidence
X
$Q_{f(t)}$
β
max
a
action / choice
$Q_{f(t+1)}$
Q(s)
predicted value
X
δ
prediction error
-
r
outcome
Confidence
α
B
Implicit decision confidence as rectified sensory evidence.
State representations
Decision confidence as uncertainty.
Value, action selection
Conscious reports of decision confidence
General decision confidence, willingness to wait
Model selection
Decision confidence across domains (perceptual, memory, ...)
Decision confidence as probability correct
Counterfactual & prospective decision confidence
Value signals: unchosen option
Explicit decision confidence
Decision confidence as certainty / uncertainty
Value signals: chosen option
Prediction errors
Confidence prediction errors
Prediction errors, cached value, action selection
Decision confidence as belief in choice accuracy
PC / LIP
SMA
DLPFC
PreC
ACC
Pul
FPC
pgACC
Pu
Cd
dMB
OFC/vmPFC
Cerebellum

Fig. 1. The diverse computational and neural substrates of confidence. A – Schematic of a simple reinforcement learning algorithm, with each model operation potentially affected by confidence indicated by a red arrow. *X*: integration, ***α***: learning rate, ***β***: stochasticity (randomness), **δ**: prediction error, *r*: outcome / reward, *a*: selected action, *Q*: value function / prediction. **B** - Neural circuits of confidence – highlighted in orange, reinforcement learning – highlighted in black, and their potential interactions – highlighted with red contour. Circuits for metacognition and confidence include subregions of the prefrontal cortex (e.g., frontal poles, DLPFC, OFC, vmPFC), parietal, precuneus, ACC, pgACC, as well as basal ganglia and thalamus (pulvinar). Areas associated with reinforcement learning are in the PFC (value estimates: vmPFC, frontal poles, SMA), in the basal ganglia (prediction errors, action selection), cerebellum (prediction errors), SMA (action selection), parietal cortex (state representations), OFC / ACC (model selection). Potential interactions can happen directly within regions codings for both confidence and reinforcement learning variables, and through deep parallel loops linking subcortical areas with neocortex. Abbreviations: DLPFC = dorsolateral prefrontal cortex, OFC = orbitofrontal cortex, vmPFC = ventromedial prefrontal cortex, ACC = anterior cingulate cortex, pgACC = pregenual ACC, SMA = supplementary motor area, FPC: frontopolar cortex, PC / LIP: parietal cortex, lateral intraparietal sulcus, PreC: precuneus, Pul: pulvinar, dMB: dopaminergic midbrain, Pu: putamen, Cd: caudate.

as variability, noise, etc).

2.3. *Stochasticity*

Relatedly, confidence may control global behavioural variables such as exploitation-exploration states. In the wild, biological organisms periodically shift from explorative states to states of exploitation, where the agent can use a set of learned rules or trajectories, to optimize returns. Algorithms using variance estimates (upper confidence bound) to regulate the exploration-exploitation trade-off have been shown to be efficient and effective in multi-armed bandit problems (Audibert et al., 2009). People too tend to use uncertainty to arbitrate between exploration and exploitation, using low confidence about value estimates to signal a switch to explorative behaviour (Boldt et al., 2019).

2.4. *Value estimates*

Standard reinforcement learning algorithms update the value function for the ‘state’ visited and the ‘action’ selected. This rests on the strong assumption that the state is implicitly or explicitly known.

However, in the real-world information about the state is often ambiguous and noisy, and the outcome of one's action may not be immediately known. It is thus often impossible to be sure about which state to update, especially when the number of possible states is high. Confidence could play a critical function (Lak et al., 2017): when the agent is confident, the value prediction update can be sharp at or around the relevant state. On the other hand, if the agent is not confident about the decision, or the state information, then the update may take a distributed form, whereby the value of neighbouring or conceptually related states may undergo partial updates. The spread of the update function towards alternative states may in itself also depend on confidence.

2.5. Prediction errors

Beyond value estimates and learning rates, confidence could also directly influence prediction errors. For example by modulating their magnitude (Lak et al., 2019), or as a step-function determining whether the prediction error should be considered at all. More complicated reinforcement learning architectures may afford additional venues for confidence computations related to prediction errors. In hierarchical models with mixture-of-experts architectures, an agent takes a global action as the weighted average, or product, of individual experts (Doya et al., 2002; Jacobs et al., 1991; Sugimoto et al., 2012). Each expert is effectively a simpler learning algorithm, tracking a specific subset of the environment or the representational space, or implementing a unique policy. Usually, prediction errors from each expert are used to weigh the experts and select the agent's global action. An exciting question here is how much confidence and metacognition can gauge these internal representations or strategies, affecting the way prediction error signals are used to select the best expert (Cortese et al., 2021).

2.6. Model selection

Cognitive control and strategic behaviour can provide further, fertile grounds to investigate higher order properties of confidence. In hierarchical fashion, multiple confidence signals tracking trial-by-trial decisions, evidence, and their outcomes could be summarised into more abstract forms of metacognition to monitor the effectiveness, or usefulness, of a certain behavioural strategy. When the reliability of the current world model – i.e., how well it predicts events in the world, or choice outcomes – goes down (Donoso et al., 2014), confidence could trigger changes in the course of action (Sarafyazd and Jazayeri, 2019). Here, reliability is strictly a variable related to the model: it is computed based on its parameters, similar to prediction errors. Confidence instead is more "general", since it can be in multiple dimensions, such as perception (e.g., about the input stimuli that are used by the model), decisions (e.g., the action selected by the model). Note the arrow of causality from reliability, to confidence, to behavioural switches remains speculative, since previous studies did not investigate confidence directly. Model reliability and confidence may evolve in parallel and affect behaviour as indicated above or, in other circumstances, orthogonally.

Are these influences of confidence (Fig. 1A) distinguishable empirically? Experimental manipulations and computational modelling of behaviour can provide insight into these different hypotheses and which parameter is most likely controlled by confidence under a given set of conditions. For example, under high volatility in outcomes' probabilities, the main effect of confidence would likely be seen on evidence accumulation or learning rate. That is, to sample information longer or to limit the horizon of relevant past experience to the last few trials. In situations of high uncertainty in the states – either because of ambiguous stimuli, or the high dimensionality of states, confidence would affect mainly evidence accumulation and value estimates. These are just two examples, a better characterization of the conditions and the uniqueness (or not) of confidence influences on reinforcement learning variables is necessary. But besides experimental manipulations, computational modelling and simulations will prove crucial to gain quantitative insight and make progress in assessing the different influences (Palminteri et al., 2017; Wilson and Collins, 2019). Computational modelling of behavioural data, possibly including neural data [as in e.g., (Leong et al., 2017)], will allow the quantification of each influence. Model comparisons will indicate which models are most likely to represent the underlying processes. Simulations across multiple candidate models (e.g., with and without confidence influence on specific learning parameters) using a range of parameter values, as well as using parameter estimates from participants' choices, will further enable the falsification of specific models that may initially appear legitimate (Palminteri et al., 2017).

Finally, it is important to note that confidence itself can be integrated as a prediction error signal in the absence of external feedback (Daniel and Pollmann, 2012; Guggenmos et al., 2016; Stolyarova et al., 2019). Thus, in perceptual learning (but other learning scenarios too, probably), confidence itself can act as the teaching signal.

3. Neural substrates at the intersection of learning and metacognition

The neural circuits that support reinforcement learning, confidence computations, and their interaction are surprisingly diverse. In reinforcement learning, research has traditionally aligned the neural architecture onto separate regions that track task and environment states – PFC and parietal cortex; prediction errors – caudate / striatum / putamen; and action selection – supplementary motor area and basal ganglia (Fig. 1B, black notes). Yet, this picture has changed as more studies examined the neural underpinnings of reinforcement learning under a wider variety of conditions. Rather than having general reinforcement learning modules that are *anatomically and functionally* segregated, the brain implements (1) single learning variables encoded in parallel across multiple regions, (2) a diversity of variables encoded by single neurons [i.e., as in mixed selectivity (Fusi et al., 2016; Rigotti et al., 2013)], (3) different variables encoded by different neurons within localized circuits. Prediction errors offer a great example: correlates have been found in almost every region of the brain, from the cerebellum (Heffley and Hull, 2019; Schlerf et al., 2012) to the frontopolar cortex (Boorman et al., 2011). Crucially, these prediction errors appear distributed according to their target behavioural dimension: sensory, motor, social, counterfactuals, etc. Moreover, even within a single task, prediction errors about different task features are encoded across a broad fronto-parietal network (Oemisch et al., 2019). Yet somewhere, most likely the ventral striatum in the basal ganglia, neurons code for general prediction errors [as an abstract signal, arising from any dimension (Schultz and Dickinson, 2000)].

Similarly, is the computation of confidence unequivocally linked to one central circuit, or is it driven by a distributed neural coding scheme (Fig. 1B, orange notes)? In humans, the PFC – and particularly the rostrolateral and dorsolateral PFC (rlPFC/DLPFC) – is a central component of the explicit metacognitive / confidence circuitry (Cortese et al., 2016; Fleming et al., 2010; Lau and Passingham, 2006; Rounis et al., 2010). Recent studies have sought to disentangle different components that contribute to the construction of confidence. DLPFC activity correlated with the reported scalar confidence, while the pregenual anterior cingulate cortex (pgACC) contributed by forming an estimate of the probability of making a correct choice (Bang and Fleming, 2018; Morales et al., 2018). Interestingly, activity in the pgACC has also been shown to correlate with the uncertainty in pain controllability (Zhang et al., 2020). This area may thus code a general certainty monitoring signal reflecting the current behavioural demand or goal: e.g., about the correctness of a choice in decision-making; about controllability in a control problem, etc.

In monkeys, neurons in the area LIP code for implicit confidence signals (Kiani et al., 2014; Kiani and Shadlen, 2009). In rodents, the ACC has been found critical for confidence in visual modality (Stolyarova et al., 2019), while the orbitofrontal cortex (OFC) in olfactory modality

(Kepecs et al., 2008; Lak et al., 2014). More recent work has shown that single OFC neurons compute abstract confidence, across sensory (olfactory and auditory) modalities (Masset et al., 2020).

Beside the neocortex, subcortical regions are also very relevant for confidence. Given the wiring patterns of these areas with brain regions involved in reinforcement learning, the subcortical coding of confidence offers an interesting path to explore the points of contact between confidence and learning (Fig. 1B). Neurons in the pulvinar (a set of thalamic nuclei) explicitly signal decision confidence: their firing rates vary linearly with the degree of confidence, and their inactivation impairs confidence behaviour without affecting performance (Jaramillo et al., 2019; Komura et al., 2013). In the basal ganglia (caudate, putamen), a confidence prediction error occurs when humans learn without feedback (Daniel and Pollmann, 2012; Guggenmos et al., 2016). Dopaminergic neurons in the neighbouring midbrain signal confidence as a belief state about choice accuracy in rats (Lak et al., 2017). Disruption of dopaminergic neurons' normal coding affects the animal's future choices, consistent with an effect of confidence on prediction errors (Lak et al., 2019).

Thus, what is the exact substrate for the interaction between confidence and reinforcement learning? For one, there are striking parallels in terms of the brain regions involved in both computation streams. Within the basal ganglia, the caudate and putamen represent prediction errors that go beyond *reward* prediction errors to also include *confidence* prediction errors. The vmPFC linearly encodes values (in reinforcement learning, expected value), but also confidence quadratically (Lebreton et al., 2015). Furthermore, top down projections from prefrontal cortex, parietal and cingulate cortices may directly pair neurons computing confidence with reinforcement learning circuits. Confidence signals could be integrated as feedback terms by different subsets of reinforcement learning-related neurons (coding prediction errors, value, action selection).

One appealing hypothesis along this line of thought is that parallel loops linking basal ganglia with cortical regions (especially the prefrontal cortex) (Haruno and Kawato, 2006; Jeon et al., 2014; Lee et al., 2020; Nakahara et al., 2001) support these interactions (Cortese et al., 2019). Anatomically, the DLPFC – implicated in explicit metacognitive reports, is richly connected to the rostral and caudal parts of the putamen and caudate, as well as the ventral striatum (Draganski et al., 2008), which represent different types of prediction errors, cached values, and commands for action selection. The parallel loops could allow the interactions to take place at multiple levels of specificity or abstraction.

These ideas bring the discussion surprisingly close to the proposal that the brain is composed of *learner* and *meta-learner* modules (Doya, 2002). Broadly speaking, the learner integrates information, associations, outcomes or behavioural trajectories that map onto limited time steps, such as a single action-outcome pair. The meta-learner instead "learns to learn": it is dedicated to extracting regularities and structure over longer time horizons, so as to tune the free parameters of the learner (Buschman and Miller, 2014; Doya, 2002; Wang et al., 2018). In the brain, the meta-learner loosely maps onto the PFC, while the learner onto the basal ganglia (Buschman and Miller, 2014; Doya, 2002). Given the ways in which confidence has been shown to affect the update of several reinforcement learning parameters (evidence accumulation, learning rates, etc), an attractive idea is that some elements of metacognitive confidence are part of a meta-learning system. Confidence could negotiate adjustments in behavioural policies via parameter updates in reinforcement learning modules. Note that this meta-learning system, although generally mapping onto the PFC, is not necessarily restricted to the PFC alone.

4. Measuring, modulating, and generating neural meta-representations

A critical question concerns how to investigate the neural representations and neural circuitry that support the interaction between confidence / metacognition and learning. In this context multivoxel neural reinforcement (also known as decoded neurofeedback) is an attractive causal approach (Shibata et al., 2011; Taschereau-Dumouchel et al., 2020), which has evolved with recent developments in neuroimaging (Feinberg et al., 2010; Haxby et al., 2014, 2011; Xu et al., 2013). Combining machine learning approaches and real-time fMRI, multivoxel neural reinforcement has high content specificity (Lubianiker et al., 2019), which enables researchers to reinforce or modify connections between two or more brain regions, specific neural representations, and even psychological processes.

In multivoxel neural reinforcement the experimenter first builds a machine learning decoder using brain activity patterns acquired with fMRI (or EEG, MEG) while participants engage in a task or are at rest. For example, while participants discriminate the movement direction of a cloud of dots. A trained decoder (e.g., that predicts left vs right motion from brain activity patterns), will be able to compute the likelihood that a new brain activity pattern represents each original class (left and right motion). Because the calculation can be done in real-time and participants need not be aware of the procedure, this approach provides a powerful means of accessing ongoing conscious or nonconscious neural representations. The output of the decoder (the probability of, say, left motion) can then be used to provide a commensurate reward and teach the brain to reinforce specific activity patterns (Fig. 2A). This way, one can induce perceptual learning (Amano et al., 2016; Shibata et al., 2011), change confidence (Cortese et al., 2017, 2016; Koizumi et al., 2017), reduce fears (Koizumi et al., 2016; Taschereau-Dumouchel et al., 2018), and – presumably – elicit more complex secondary associations. Importantly, the vast majority of participants do not have conscious access to the information represented in their momentary brain activity patterns during decoded neurofeedback (Shibata et al., 2019).

Richer designs can be considered in which task conditions are directly determined by high-dimensional multivoxel patterns of activity (Chew et al., 2019; Cortese et al., 2020). Thus, this kind of paradigm provides a useful way to study the mechanisms by which the brain resolves dimensionality issues with limited amounts of training data (the "curse of dimensionality"). In a recent study by Cortese et al., an action (e.g., choosing option A) was paired with a particular latent brain state (e.g., a nonconscious representation of leftward motion) (Cortese et al., 2020). Among a huge number of possible ongoing patterns of neural activity (with additional noise since measured indirectly through fMRI), the brain had to learn to selectively associate the output of a locally restricted neural population (e.g., within the visual cortex), with an action, through trial-and-error. The problem faced by the brain is effectively equivalent to searching a needle in a haystack. Yet, participants displayed learning evidence.

Crucially, Cortese and colleagues found that perceptual confidence indexed reinforcement learning at multiple levels (Cortese et al., 2020). Across participants, initial metacognitive sensitivity (Maniscalco and Lau, 2012) – defined as the ability to discriminate between correct and incorrect responses – predicted subsequent learning performance in the high-dimensional reinforcement learning task. Second, decision confidence was positively correlated with reinforced choice accuracy, and negatively correlated with the trial-by-trial magnitude of prediction errors obtained from reinforcement learning modeling. On trials in which participants' confidence was lowest, they exhibited large learning uncertainty (Fig. 2B). That is, prediction errors had large magnitude indicating large mismatches between expectations and actual outcomes. This relationship was supported by the neural coupling between the DLPFC and the basal ganglia. Specifically, with learning, DLPFC activity patterns representing perceptual confidence became increasingly associated with basal ganglia activity patterns representing prediction error magnitude (Fig. 2C). Moreover, the functional connectivity measured at rest between the basal ganglia and a subregion of the DLPFC also increased as learning progressed, further supporting the finding of an association between metacognition and reinforcement learning at the neural level. Taken together, these results support a model of the brain

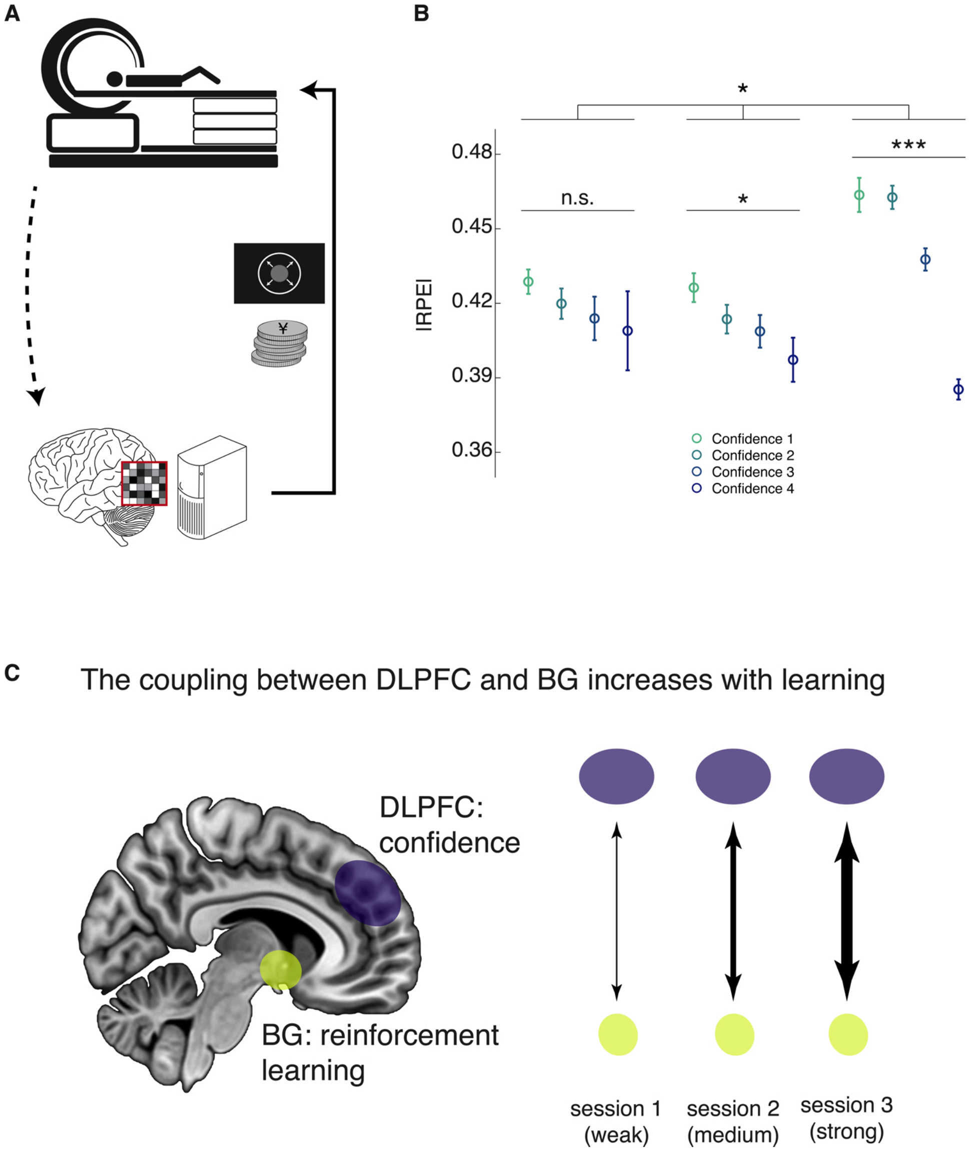

Fig. 2. Multivoxel neural reinforcement and closed-loop experiments for confidence and learning. A - The decoded neurofeedback protocol. A participant's brain activity is measured with fMRI, and the data is used in real-time to compute a score with a machine learning decoder. This score signals the likelihood that the current activity pattern represents the target, such as an object, a task feature, or a psychological variable. In the neurofeedback experiment, the score is fed back to participants in the form of a visual cue (a circle). Participants are instructed that the larger the feedback circle, the larger the monetary reward. The same setting can be utilized to define task conditions. That is, instead of using the decoder score to provide visual feedback, the score can be used to determine the contingency for an optimal choice, or to modify a visual stimulus presented on the screen, or to determine the task difficulty, in real-time. Image modified from (Cortese et al., 2021). **B** - confidence was associated with prediction error magnitude in a study where participants had to learn a high-dimensional mapping between patterns of brain activity and a 1-d (2 options) action space. That is, the contingency for the optimal choice was defined by a brain activity pattern. The results show that the higher the confidence, the smaller the learning uncertainty in the form of smaller prediction errors. * $p<0.05$, *** $p<0.001$. Image modified from (Cortese et al., 2020). **C** - In parallel with confidence - prediction error magnitude correlation at the behavioural and computational levels, the same study found increased neural coupling between the DLPFC and the basal ganglia. This coupling was specific for the information represented in these areas: confidence in DLPFC, and uncertainty (prediction error magnitude) in the basal ganglia, and it increased with learning. Image modified from (Cortese et al., 2020).

in which metacognition (confidence) is causally important for reinforcement learning. Although the demonstration was not direct because it was based on observational data, analysis on time lag effects, as well as additional analyses ruling out alternative interpretations, make it highly plausible that metacognition plays a causal function in reinforcement learning.

While this study established the role of metacognition in high-dimensional reinforcement learning problems, the actual mechanism remains essentially unclear. Parallel loops between the basal ganglia and the prefrontal cortex (and probably elsewhere in the brain too) provide a strong candidate neural substrate, but it remains unclear what is the actual implementation at the neuronal (and representational) levels.

Taking a more speculative stance, these results may suggest that metacognition accelerates reinforcement learning processes by generating low-dimensional meta-representations. When confidence is computed about a choice (as in this experiment, but it could be about something else such as a memory, or the sensory stimulus itself), the brain may *re*-represent the content of the primary (first-order) representation as a meta-representation. That is, suppose the choice is about motion direction, then there should be a primary motion representation, in area V5 of the visual cortex. By computing confidence about the relevant choice, the brain could re-represent this motion information in the prefrontal cortex, as a meta-representation. Such meta-representations are void of irrelevant information, and are construed in a low-dimensional space. If reinforcement learning operates on these low-dimensional folds, it would be faster, bypassing the bottlenecks that arise in high-dimensional sensory space.

The use of closed-loop neuroimaging and machine learning decoders provides a unique opportunity to quantify the function of neural representations (conscious or nonconscious), meta-representations, in metacognition and learning.

5. Discussion, perspectives, and the many open questions

This paper aimed to provide an overview of the burgeoning area of research on metacognition and learning, at the intersection of cognitive sciences, psychology and computational neuroscience. It is becoming increasingly clear that confidence can affect learning processes at various levels: by controlling how evidence is accumulated, by directly biasing value estimates, by modulating learning rates, by affecting prediction errors and their integration, and by changing the amount of exploration drive for the selection of new strategies. From a more speculative viewpoint, metacognition could also accelerate learning by generating meta-representations, essential building blocks of summarised abstract states in the brain.

Do we now know where in the brain confidence signals originate? In the quest to characterize the neural and computational origins of confidence, invariably there will be differences across studies (as discussed in the section about neural correlates, Fig. 1B). Similar to other fields in decision-making research (Wallis, 2011), some of the differences in the literature might arise because of complex functional homologies between areas across species (humans, non-human primates, rodents, …). This may be particularly true between rodents and non-human primates or humans (Wise, 2008), as the latter two display high anatomical and functional homology in terms of brain areas related to reward-guided learning and decisions (Neubert et al., 2015). The primary decision to record from a given area might have an impact too – we might be recording from the ACC, but neurons in the OFC are perhaps also related to the same experimental variable(s). Yet, differences probably also arise from the type of confidence measured (reflecting statistical confidence or a metacognitive evaluation), or the tasks used. As mentioned earlier, humans are often asked to report confidence as an explicit judgement on a linear scale, while in other animals confidence is measured as the willingness to wait for a reward or choosing an opt-out safe option. Although all these tasks clearly measure confidence signals, the underlying computations and the behavioural endpoint can subtly differ. Future work could capitalize on the use of closely matched tasks and conditions in different species (Odegaard et al., 2018; Stolyarova et al., 2019) to disentangle these issues.

From a more conceptual angle, can we really talk about meta-representations (re-representations of ongoing sensory or memory representations)? In a way, from a purely mechanistic viewpoint, meta-representations are here equivalent to abstractions: simplified representations, or schematics, of more complex information (Fleming, 2020). Crucially, these representations should be found in the prefrontal cortex, as a relevant higher-order representation that directly stems from an inner monitoring system (Brown et al., 2019). Because reinforcement learning processes would need immediate access to these meta-representations to be efficient, there appears here an interesting overlap – if not functionally, at least in anatomical terms – on the neural structure of parallel loops between the basal ganglia and the neocortex discussed earlier. These loops connect reinforcement learning with confidence and metacognitive signals, but also with meta-representations and other high cognitive functions. A speculation here is that confidence is involved in the construction of meta-representations. But could it be rather that confidence is involved in the way reinforcement learning selects and operates on meta-representations? An exciting avenue of research here is the possibility that the interaction between metacognition and reinforcement learning, over longer time scales, results in the construction of schemas and cognitive maps (i.e., internal world models) through assembly of meta-representations.

Relatedly, once multiple internal models have been constructed, confidence could still be part of the selection / arbitration mechanism favouring one rather than the other strategy. Considering this arbitration mechanism, what kind of representational architecture is the most ‘efficient’? That is, should the brain track multiple strategies simultaneously - as in a mixture-of-experts architecture, or should it represent only the best strategy at any point in time, keeping the rest latent? In both cases confidence can help in selecting the most relevant or useful strategy, but each approach has benefits and limitations. A mixture-of-experts architecture means that a more complex problem can be broken down into simpler components, and is very useful when data is sparse and there is parallel computational power, since with a single data point multiple internal models can be updated simultaneously. It is also computationally costly and has the downside of computing behavioural trajectories that may never be used. Conversely, a hypothesis-testing regime, where the best strategy or internal model is actively represented, probably allows faster computing when resources are limited, or the environment is noisy but also incorporates exploitable statistical regularities. Unfortunately, this also means that the agent could be slower to update its course of action when environmental conditions change suddenly, because the alternatives are not all up-to-date. What is implemented by the brain? Evidence to date indicates both solutions are plausible (Badre and Frank, 2012; Donoso et al., 2014; Frank and Badre, 2012). The brain may capitalize on both types of computational approaches, separately or in hybrid fashion depending on the circumstances.

While this review covered at length the points of contact between metacognition and reinforcement learning from the viewpoint of confidence affecting learning parameters, it is important to stress here that the interaction is most likely bidirectional. Confidence is biased by previous prediction errors (Cortese et al., 2020), context and rewards (Lebreton et al., 2019, 2018), choice history (Benwell et al., 2019), post-decisional information (Navajas et al., 2016), priors (Locke et al., 2020) and learning itself (Chen et al., 2019).

One interesting point that deserves deeper investigation in the future is the intersection between inductive biases – especially acquired ones – and metacognition in the context of learning. Inductive biases are the set of assumptions, at the architecture or functional level, that guide the way an agent will apply knowledge to novel situations. In machine learning, and especially in deep learning algorithms, integrating high-level inductive biases is now viewed as a critical step to build AI systems that enjoy similar levels of flexibility and generalization as humans (Goyal and Bengio, 2020; Tenenbaum et al., 2011). A compelling example of inductive bias is compositionality, whereby simpler building blocks are combined at will to create more complex structure, which allow agents to make strong inference for generalization with small amounts of data. When humans learn new concepts, compositionality is a ubiquitous strategy (Goodman et al., 2008; Kemp, 2012). Although inductive biases do not need metacognition to operate (Tenenbaum et al., 2011), recent work in cognitive neuroscience has shown how humans have metacognitive access into their own concept learning process (Stojic et al., 2018), opening a first link between inductive biases and metacognition.

Intriguingly, we often discuss humans in terms of intelligence, extreme generalization abilities and flexible behaviours; but humans are also beguiled by habits, irrational choices, and in general behaviours that appear rather inflexible. A strong example in this context is confirmation bias, whereby we seek or interpret evidence in ways that are partial to our existing beliefs or expectations (Nickerson, 1998). Why this tension? Perhaps some answers are to be found in the specifics of metacognition and self-monitoring abilities. Here the link develops over two levels: confidence itself relying more on the information that supports the decision made (Michel and Peters, 2020; Zylberberg et al., 2012), or the latest choices (Talluri et al., 2018), as well as a more general tendency of individuals with alterations in metacognition to display high confirmation bias (Rollwage et al., 2018). In fact, when confirmation bias is a built-in part of a metacognitive agent, it can become adaptive (Rollwage and Fleming, 2021). That is, an agent showing high levels of self-awareness and confirmation bias performs better than an unbiased agent. Intuitively, such an agent should better distinguish situations in which errors are more likely – confidence would be low – and situations in which responses are mostly correct (high confidence), thus allowing selective information processing. In this context, it would be interesting to uncover the interactions between neural systems dedicated to learning new information, skills or behaviours, those involved in metacognition and self-monitoring, and those tailored to exploit a certain situation or knowledge. This way we may understand how the brain actively *ignores* information (especially when only some information is central to the problem or behavioural demand at hand), and how confirmation bias can be adaptive or maladaptive for high level strategic behaviour.

In sum, the goal of this manuscript was to shine a light on the fascinating computational properties of metacognition and confidence in learning. Particularly, highlighting points of discussion that may be useful stepping stones to further study the nature of adaptive behaviours, efficient learning, and intelligence. Findings from these lines of research will have profound implications for our appreciation of high level brain functions, for understanding maladaptive learning processes, but also for the development of novel artificial agents. The decoded neurofeedback approach will further allow novel experiments that reach beyond the standard logic of neuroscience and psychology experiments, by using the brain as its own canvas.

Declaration of Competing Interest

The author declares no competing interests.

Acknowledgements

This work was supported by the Ikegaya Brain-AI Hybrid ERATO grant (JPMJER1801) from the Japan Science and Technology Agency., and by the Innovative Science and Technology Initiative for Security grant (JPJ004596) from ATLA (Japan).

References

Amano, K., Shibata, K., Kawato, M., Sasaki, Y., Watanabe, T., 2016. Learning to associate orientation with color in early visual areas by associative decoded fMRI neurofeedback. Curr. Biol. 26, 1861–1866.

Audibert, J.-Y., Munos, R., Szepesvári, C., 2009. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. Theor. Comput. Sci. 410, 1876–1902.

Badre, D., Frank, M., 2012. Mechanisms of hierarchical reinforcement learning in cortico–striatal circuits 2: evidence from fMRI. Cereb. Cortex 22, 527–536.

Balsdon, T., Wyart, V., Mamassian, P., 2020. Confidence controls perceptual evidence accumulation. Nat. Commun. 11, 1–11.

Bang, D., Fleming, S.M., 2018. Distinct encoding of decision confidence in human medial prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1800795115.

Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K. L., Kurth-Nelson, Z., 2018. What is a cognitive map? Organizing knowledge for flexible behavior. Neuron 100, 490–509.

Bengio, Y., 2017. The Consciousness Prior. arXiv [cs.LG].

Benwell, C.S.Y., Beyer, R., Wallington, F., Ince, R.A.A., 2019. History biases reveal novel dissociations between perceptual and metacognitive decision-making. bioRxiv. https://doi.org/10.1101/737999.

Boldt, A., Blundell, C., De Martino, B., 2019. Confidence modulates exploration and exploitation in value-based learning. Neurosci. Conscious. https://doi.org/10.1093/nc/niz004.

Boorman, E., Behrens, T., Rushworth, M., 2011. Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. PLoS Biol. 9, e1001093.

Botvinick, M., Ritter, S., Wang, J.X., Kurth-Nelson, Z., Blundell, C., Hassabis, D., 2019. Reinforcement learning, fast and slow. Trends Cogn. Sci. https://doi.org/10.1016/j.tics.2019.02.006.

Brosnan, M.B., Sabaroedin, K., Silk, T., Genc, S., Newman, D.P., Loughnane, G.M., Fornito, A., O'Connell, R.G., Bellgrove, M.A., 2020. Evidence accumulation during perceptual decisions in humans varies as a function of dorsal frontoparietal organization. Nat. Hum. Behav. 4, 844–855.

Brown, R., Lau, H., LeDoux, J.E., 2019. Understanding the higher-order approach to consciousness. Trends Cogn. Sci. 23, 754–768.

Buschman, T.J., Miller, E.K., 2014. Goal-direction and top-down control. Philos. Trans. R. Soc. Lond. B Biol. Sci. 369, 20130471–20130471.

Chen, B., Mundy, M., Tsuchiya, N., 2019. Metacognitive accuracy improves with the perceptual learning of a low- but not high-level face property. Front. Psychol. 10, 1712.

Chew, B., Hauser, T.U., Papoutsi, M., Magerkurth, J., Dolan, R.J., Rutledge, R.B., 2019. Endogenous fluctuations in the dopaminergic midbrain drive behavioral choice variability. Proc. Natl. Acad. Sci. U. S. A., 201900872

Cortese, A., Amano, K., Koizumi, A., Kawato, M., Lau, H., 2016. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat. Commun. 7, 13669.

Cortese, A., Amano, K., Koizumi, A., Lau, H., Kawato, M., 2017. Decoded fMRI neurofeedback can induce bidirectional confidence changes within single participants. NeuroImage 149, 323–337.

Cortese, A., De Martino, B., Kawato, M., 2019. The neural and cognitive architecture for learning from a small sample. Curr. Opin. Neurobiol. 55, 133–141.

Cortese, A., Lau, H., Kawato, M., 2020. Unconscious reinforcement learning of hidden brain states supported by confidence. Nat. Commun. 1–14.

Cortese, A., Yamamoto, A., Hashemzadeh, M., Sepulveda, P., Kawato, M., De Martino, B., 2021. Value signals guide abstraction during learning. Elife 10. https://doi.org/10.7554/eLife.68943.

Daniel, R., Pollmann, S., 2012. Striatal activations signal prediction errors on confidence in the absence of external feedback. Neuroimage 59, 3457–3467.

Dayan, P., Niv, Y., 2008. Reinforcement learning: the good, the bad and the ugly. Curr. Opin. Neurobiol. 18, 185–196.

De Martino, B., Fleming, S., Garrett, N., Dolan, R., 2012. Confidence in value-based choice. Nat. Neurosci. 16, 105–110.

Dehaene, S., Lau, H., Kouider, S., 2017. What is consciousness, and could machines have it? Science 358, 486–492.

Donoso, M., Collins, A.G.E., Koechlin, E., 2014. Foundations of human reasoning in the prefrontal cortex. Science 344, 1481–1486.

Doya, K., 2002. Metalearning and neuromodulation. Neural Netw. 15, 495–506.

Doya, K., 2007. Reinforcement learning: computational theory and biological mechanisms. HFSP J. 1, 30–40.

Doya, K., Samejima, K., Katagiri, K.-I., Kawato, M., 2002. Multiple model-based reinforcement learning. Neural Comput. 14, 1347–1369.

Draganski, B., Kherif, F., Klöppel, S., Cook, P., Alexander, D., Parker, G., Deichmann, R., Ashburner, J., Frackowiak, R., 2008. Evidence for segregated and integrative connectivity patterns in the human basal ganglia. J. Neurosci. 28, 7143–7152.

Drugowitsch, J., Mendonça, A.G., Mainen, Z.F., Pouget, A., 2019. Learning optimal decisions with confidence. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1906787116.

Feinberg, D.A., Moeller, S., Smith, S.M., Auerbach, E., Ramanna, S., Gunther, M., Glasser, M.F., Miller, K.L., Ugurbil, K., Yacoub, E., 2010. Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. PLoS One 5, e15710.

Fleming, S.M., 2020. Awareness as inference in a higher-order state space. Neurosci. Conscious. 6 (1), niz020.

Fleming, S., Lau, H., 2014. How to measure metacognition. Front. Hum. Neurosci. 8 https://doi.org/10.3389/fnhum.2014.00443.

Fleming, S., Weil, R., Nagy, Z., Dolan, R., Rees, G., 2010. Relating introspective accuracy to individual differences in brain structure. Science 329, 15411543.

Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: computation, biology and function. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 1280–1286.

Folke, T., Jacobsen, C., Fleming, S.M., De Martino, B., 2016. Explicit representation of confidence informs future value-based decisions. Nat. Hum. Behav. 1, 0002.

Frank, M.J., Badre, D., 2012. Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: computational analysis. Cereb. Cortex 22, 509–526.

Fusi, S., Miller, E.K., Rigotti, M., 2016. Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. 37, 66–74.

Goodman, N.D., Tenenbaum, J.B., Feldman, J., Griffiths, T.L., 2008. A rational analysis of rule-based concept learning. Cogn. Sci. 32, 108–154.

Goyal, A., Bengio, Y., 2020. Inductive Biases for Deep Learning of Higher-Level Cognition. arXiv [cs.LG].

Guggenmos, M., Wilbertz, G., Hebart, M., Sterzer, P., 2016. Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. eLife 5. https://doi.org/10.7554/eLife.13388.

Haruno, M., Kawato, M., 2006. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. Neural Netw. 19, 1242–1254.

Haxby, J., Guntupalli, J., Connolly, A., Halchenko, Y., Conroy, B., Gobbini, M., Hanke, M., Ramadge, P., 2011. A common, high-dimensional model of the representational space in human ventral temporal cortex. Neuron 72. https://doi.org/10.1016/j.neuron.2011.08.026.

Haxby, J., Connolly, A., Guntupalli, J., 2014. Decoding neural representational spaces using multivariate pattern analysis. Annu. Rev. Neurosci. 37, 435–456.

Heffley, W., Hull, C., 2019. Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. Elife 8. https://doi.org/10.7554/eLife.46764.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., 1991. Adaptive mixtures of local experts. Neural Comput. 3, 79–87.

Jaramillo, J., Mejias, J.F., Wang, X.-J., 2019. Engagement of pulvino-cortical feedforward and feedback pathways in cognitive computations. Neuron 101, 321–336.e9.

Jeon, H.-A., Anwander, A., Friederici, A., 2014. Functional network mirrored in the prefrontal cortex, caudate nucleus, and thalamus: high-resolution functional imaging and structural connectivity. J. Neurosci. 34, 9202–9212.

Kawato, M., Furukawa, K., Suzuki, R., 1987. A hierarchical neural-network model for control and learning of voluntary movement. Biol. Cybern. 57, 169–185.

Kemp, C., 2012. Exploring the conceptual universe. Psychol. Rev. 119, 685–722.

Kepecs, A., Mainen, Z., 2012. A computational framework for the study of confidence in humans and animals. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 1322–1337.

Kepecs, A., Uchida, N., Zariwala, H., Mainen, Z., 2008. Neural correlates, computation and behavioural impact of decision confidence. Nature 455, 227–231.

Kiani, R., Shadlen, M.N., 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324, 759–764.

Kiani, R., Corthell, L., Shadlen, M., 2014. Choice certainty is informed by both evidence and decision time. Neuron 84. https://doi.org/10.1016/j.neuron.2014.12.015.

Koizumi, A., Amano, K., Cortese, A., Shibata, K., Yoshida, W., Seymour, B., Kawato, M., Lau, H., 2016. Fear reduction without fear through reinforcement of neural activity that bypasses conscious exposure. Nat. Hum. Behav. 1, 0006.

Koizumi, A., Cortese, A., Amano, K., Kawato, M., Lau, H., 2017. Modulation of metacognition with decoded neurofeedback. Brain Nerve 69, 1427–1432.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., Miyamoto, A., 2013. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nat. Neurosci. 16, 749–755.

Lak, A., Costa, G., Romberg, E., Koulakov, A., Mainen, Z., Kepecs, A., 2014. Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron. https://doi.org/10.1016/j.neuron.2014.08.039.

Lak, A., Nomoto, K., Keramati, M., Sakagami, M., Kepecs, A., 2017. Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. Curr. Biol. 27, 821–832.

Lak, A., Okun, M., Moss, M.M., Gurnani, H., Farrell, K., Wells, M.J., Reddy, C.B., Kepecs, A., Harris, K.D., Carandini, M., 2019. Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. Neuron. https://doi.org/10.1016/j.neuron.2019.11.018.

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., E, A., Donner, T.H., Carandini, M., Tonegawa, S., Uchida, N., Kepecs, A., 2020. Reinforcement biases subsequent perceptual decisions when confidence is low: a widespread behavioral phenomenon. eLife 9, e49834.

Lake, B., Salakhutdinov, R., Tenenbaum, J., 2015. Human-level concept learning through probabilistic program induction. Science 350, 1332–1338.

Lau, H., Passingham, R., 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc. Natl. Acad. Sci. 103, 18763–18768.

Lau, H., Rosenthal, D., 2011. Empirical support for higher-order theories of conscious awareness. Trends Cogn. Sci. 15, 365373.

Lebreton, M., Abitbol, R., Daunizeau, J., Pessiglione, M., 2015. Automatic integration of confidence in the brain valuation signal. Nat. Neurosci. 18, 1159–1167.

Lebreton, M., Langdon, S., Slieker, M.J., Nooitgedacht, J.S., Goudriaan, A.E., Denys, D., van Holst, R.J., Luigjes, J., 2018. Two sides of the same coin: monetary incentives concurrently improve and bias confidence judgments. Sci. Adv. 4, eaaq0668.

Lebreton, M., Bacily, K., Palminteri, S., Engelmann, J.B., 2019. Contextual influence on confidence judgments in human reinforcement learning. PLoS Comput. Biol. 15, e1006973.

Lee, J., Wang, W., Sabatini, B.L., 2020. Anatomically segregated basal ganglia pathways allow parallel behavioral modulation. Nat. Neurosci. https://doi.org/10.1038/s41593-020-00712-5.

Leong, Y., Radulescu, A., Daniel, R., Vivian, D., Niv, Y., 2017. Dynamic interaction between reinforcement learning and attention in multidimensional environments. Neuron 93, 451–463.

Lim, K., Wang, W., Merfeld, D.M., 2020. Frontal scalp potentials foretell perceptual choice confidence. J. Neurophysiol. 123, 1566–1577.

Locke, S.M., Gaffin-Cahn, E., Hosseinizaveh, N., Mamassian, P., Landy, M.S., 2020. Priors and payoffs in confidence judgments. Atten. Percept. Psychophys. 82, 3158–3175.

Lubianiker, N., Goldway, N., Fruchtman-Steinbok, T., Paret, C., Keynan, J.N., Singer, N., Cohen, A., Kadosh, K.C., Linden, D.E.J., Hendler, T., 2019. Process-based framework for precise neuromodulation. Nat. Hum. Behav. 3, 436–445.

Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422430.

Maniscalco, B., Lau, H., 2016. The signal processing architecture underlying subjective reports of sensory awareness. Neurosci. Conscious. https://doi.org/10.1093/nc/niw002.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S.H., Basso, M.A., Lau, H., Peters, M.A.K., 2021. Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. PLoS Comput. Biol. 17, e1008779.

Masset, P., Ott, T., Lak, A., Hirokawa, J., Kepecs, A., 2020. Behavior- and modality-general representation of confidence in orbitofrontal cortex. Cell. https://doi.org/10.1016/j.cell.2020.05.022.

McCurdy, L.Y., Maniscalco, B., Metcalfe, J., Liu, K.Y., de Lange, F.P., Lau, H., 2013. Anatomical coupling between distinct metacognitive systems for memory and visual perception. J. Neurosci. 33, 1897–1906.

Metcalfe, J., 2008. Evolution of metacognition. Handbook of Metamemory and Memory.

Metcalfe, J., Son, L.K., 2012. Anoetic, noetic, and autonoetic metacognition. Found. Metacogn. https://doi.org/10.1093/acprof:oso/9780199646739.003.0019.

Meyniel, F., Sigman, M., Mainen, Z.F., 2015. Confidence as Bayesian probability: from neural origins to behavior. Neuron 88, 78–92.

Michel, M., Peters, M.A.K., 2020. Confirmation bias without rhyme or reason. Synthese. https://doi.org/10.1007/s11229-020-02910-x.

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., Miyashita, Y., 2017. Causal neural network of metamemory for retrospection in primates. Science 355, 188–193.

Miyazaki, K., Miyazaki, K.W., Yamanaka, A., Tokuda, T., Tanaka, K.F., Doya, K., 2018. Reward probability and timing uncertainty alter the effect of dorsal raphe serotonin neurons on patience. Nat. Commun. 9, 2048.

Morales, J., Lau, H., Fleming, S., 2018. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. J. Neurosci. 2360–2317.

Morales, J., Odegaard, B., Maniscalco, B., 2019. The neural substrates of conscious perception without performance confounds. Anthol. Neurosci. Philos.

Nakahara, H., Doya, K., Hikosaka, O., 2001. Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences - a computational approach. J. Cogn. Neurosci. 13, 626–647.

Nassar, M.R., Wilson, R.C., Heasly, B., Gold, J.I., 2010. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. J. Neurosci. 30, 12366–12378.

Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat. Neurosci. 15, 1040–1046.

Navajas, J., Bahrami, B., Latham, P., 2016. Post-decisional accounts of biases in confidence. Curr. Opin. Behav. Sci. 11, 55–60.

Neubert, F.-X., Mars, R.B., Sallet, J., Rushworth, M.F.S., 2015. Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. Proc. Natl. Acad. Sci. U. S. A. 112, E2695–704.

Nickerson, R.S., 1998. Confirmation bias: a ubiquitous phenomenon in many guises. Rev. Gen. Psychol. 2, 175–220.

Niv, Y., 2019. Learning task-state representations. Nat. Neurosci. 22, 1544–1553.

Odegaard, B., Grimaldi, P., Cho, S.H., Peters, M.A.K., Lau, H., Basso, M.A., 2018. Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. Proc. Natl. Acad. Sci. U. S. A. 115, E1588–E1597.

Oemisch, M., Westendorff, S., Azimi, M., Hassani, S.A., Ardid, S., Tiesinga, P., Womelsdorf, T., 2019. Feature-specific prediction errors and surprise across macaque fronto-striatal circuits. Nat. Commun. 10, 176.

Palminteri, S., Wyart, V., Koechlin, E., 2017. The importance of falsification in computational cognitive modeling. Trends Cogn. Sci. 21, 425–433.

Persaud, N., Peter, M., Cowey, A., 2007. Post-decision wagering objectively measures awareness. Nat. Neurosci. 10, 257–261.

Peters, M.A.K., Lau, H., 2015. Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. Elife 4, e09651.

Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic quantities for different goals. Nat. Neurosci. 19, 366–374.

Rahnev, D., Desender, K., Lee, A.L.F., Adler, W.T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L.Y., Balcı, F., Bang, J.W., Bègue, I., Birney, D.P., Brady, T.F., Calder-Travis, J., Chetverikov, A., Clark, T.K., Davranche, K., Denison, R.N., Dildine, T.C., Double, K.S., Duyan, Y.A., Faivre, N., Fallow, K., Filevich, E., Gajdos, T., Gallagher, R.M., de Gardelle, V., Gherman, S., Haddara, N., Hainguerlot, M., Hsu, T.-Y., Hu, X., Iturrate, I., Jaquiery, M., Kantner, J., Koculak, M., Konishi, M., Koß, C., Kvam, P.D., Kwok, S.C., Lebreton, M., Lempert, K. M., Ming Lo, C., Luo, L., Maniscalco, B., Martin, A., Massoni, S., Matthews, J., Mazancieux, A., Merfeld, D.M., O'Hora, D., Palser, E.R., Paulewicz, B., Pereira, M., Peters, C., Philiastides, M.G., Pfuhl, G., Prieto, F., Rausch, M., Recht, S., Reyes, G., Rouault, M., Sackur, J., Sadeghi, S., Samaha, J., Seow, T.X.F., Shekhar, M., Sherman, M.T., Siedlecka, M., Skóra, Z., Song, C., Soto, D., Sun, S., van Boxtel, J.J.A., Wang, S., Weidemann, C.T., Weindel, G., Wierzchoń, M., Xu, X., Ye, Q., Yeon, J., Zou, F., Zylberberg, A., 2020. The confidence database. Nat. Hum. Behav. https://doi.org/10.1038/s41562-019-0813-1.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., Fusi, S., 2013. The importance of mixed selectivity in complex cognitive tasks. Nature 497, 585–590.

Rollwage, M., Fleming, S.M., 2021. Confirmation bias is adaptive when coupled with efficient metacognition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200131.

Rollwage, M., Dolan, R.J., Fleming, S.M., 2018. Metacognitive failure as a feature of those holding radical beliefs. Curr. Biol. 28, 4014–4021.e8.

Rounis, E., Maniscalco, B., Rothwell, J., Passingham, R., Lau, H., 2010. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn. Neurosci. 1, 165175.

Sanders, J.I., Hangya, B., Kepecs, A., 2016. Signatures of a statistical computation in the human sense of confidence. Neuron 90, 499–506.

Sarafyazd, M., Jazayeri, M., 2019. Hierarchical reasoning by neural circuits in the frontal cortex. Science 364. https://doi.org/10.1126/science.aav8911.

Schlerf, J., Ivry, R.B., Diedrichsen, J., 2012. Encoding of sensory prediction errors in the human cerebellum. J. Neurosci. 32, 4913–4922.

Schultz, W., Dickinson, A., 2000. Neuronal coding of prediction errors. Annu. Rev. Neurosci. 23, 473–500.

Shibata, K., Watanabe, T., Sasaki, Y., Kawato, M., 2011. Perceptual learning incepted by decoded fMRI neurofeedback without stimulus presentation. Science 334, 1413–1415.

Shibata, K., Lisi, G., Cortese, A., Watanabe, T., Sasaki, Y., Kawato, M., 2019. Toward a comprehensive understanding of the neural mechanisms of decoded neurofeedback. NeuroImage 188, 539–556.

Sorokin, I., Seleznev, A., Pavlov, M., Fedorov, A., Ignateva, A., 2015. Deep Attention Recurrent Q-Network. arXiv [cs.LG].

Stojic, H., Eldar, E., Bassam, H., Dayan, P., Dolan, R.J., 2018. Are you sure about that? On the origins of confidence in concept learning. 2018 Conference on Cognitive Computational Neuroscience. https://doi.org/10.32470/ccn.2018.1197-0.

Stolyarova, A., Rakhshan, M., Hart, E.E., O'Dell, T.J., Peters, M.A.K., Lau, H., Soltani, A., Izquierdo, A., 2019. Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat. Commun. 10, 4704.

Sugimoto, N., Haruno, M., Doya, K., Kawato, M., 2012. MOSAIC for multiple-reward environments. Neural Comput. 24, 577–606.

Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press.

Talluri, B.C., Urai, A.E., Tsetsos, K., Usher, M., Donner, T.H., 2018. Confirmation bias through selective overweighting of choice-consistent evidence. Curr. Biol. 28, 3128–3135.e8.

Taschereau-Dumouchel, V., Cortese, A., Chiba, T., Knotts, J., Kawato, M., Lau, H., 2018. Towards an unconscious neural reinforcement intervention for common fears. Proc. Natl. Acad. Sci. 115, 201721572.

Taschereau-Dumouchel, V., Cortese, A., Lau, H., Kawato, M., 2020. Conducting decoded neurofeedback studies. Soc. Cogn. Affect. Neurosci. https://doi.org/10.1093/scan/nsaa063.

Tenenbaum, J., Kemp, C., Griffiths, T., Goodman, N., 2011. How to grow a mind: statistics, structure, and abstraction. Science 331, 1279–1285.

Wallis, J., 2011. Cross-species studies of orbitofrontal cortex and value-based decision-making. Nat. Neurosci. 15, 13–19.

Wang, J.X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J.Z., Hassabis, D., Botvinick, M., 2018. Prefrontal cortex as a meta-reinforcement learning system. Nat. Neurosci. https://doi.org/10.1038/s41593-018-0147-8.

Watkins, C.J.C., Dayan, P., 1992. Q-learning. Mach. Learn. 8, 279–292.

Wilson, R.C., Collins, A.G., 2019. Ten simple rules for the computational modeling of behavioral data. Elife 8. https://doi.org/10.7554/eLife.49547.

Wise, S.P., 2008. Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. 31, 599–608.

Xu, J., Moeller, S., Auerbach, E.J., Strupp, J., Smith, S.M., Feinberg, D.A., Yacoub, E., Uğurbil, K., 2013. Evaluation of slice accelerations using multiband echo planar imaging at 3 T. Neuroimage 83, 991–1001.

Zhang, S., Yoshida, W., Mano, H., Yanagisawa, T., Mancini, F., Shibata, K., Kawato, M., Seymour, B., 2020. Pain control by co-adaptive learning in a brain-machine interface. Curr. Biol. https://doi.org/10.1016/j.cub.2020.07.066.

Zylberberg, A., Barttfeld, P., Sigman, M., 2012. The construction of confidence in a perceptual decision. Front. Integr. Neurosci. 6 https://doi.org/10.3389/fnint.2012.00079.