nature reviews psychology https://doi.org/10.1038/s44159-024-00304-1

Perspective Check for updates

Understanding the development of reward learning through the lens of meta-learning

Kate Nussenbaum[1,2] & Catherine A. Hartley[1,3]

Abstract

Determining how environments shape how people learn is central to understanding individual differences in goal-directed behaviour. Studies of the effects of early-life adversity on reward learning have revealed that the environments that infants and children experience exert lasting influences on reward-guided behaviour. However, the varied findings from this research are difficult to reconcile under a unified computational account. Studies of adaptive reinforcement learning have demonstrated that learning algorithms and parameters dynamically adapt to support reward-guided behaviour in varied contexts, but this body of research has largely focused on learning that proceeds within the short timeframes of experimental tasks. In this Perspective, we argue that, to understand how the structure of experienced environments shapes reward learning across development, computational accounts of the effects of environmental statistics on reinforcement learning need to be extended to encompass learning across multiple nested timescales of experience. To this end, we consider the development of reward learning through the lens of meta-learning models, in particular meta-reinforcement learning. This computational formalization can inspire new hypotheses and methods for empirical research to understand how features of experienced environments give rise to individual differences in learning and adaptive behaviour across development.

Sections

[1]Department of Psychology, New York University, New York, NY, USA. [2]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA. [3]Center for Neural Science, New York University, New York, NY, USA. e-mail: katenuss@princeton.edu; cate@nyu.edu

Introduction

A central aim of developmental science is to understand how experience shapes the emergence and organization of goal-directed behaviour. There has been widespread investigation of how differences in the early-life environments of children relate to individual differences in the development of varied neurocognitive processes, including perceptual processing[1,2], language learning[3], motor skill acquisition[4,5] and 'higher-level' executive functioning[6,7]. Across these investigations, a common theme has emerged: salient features of early environments shape not just what information is available for learning but also how that information is processed and used to guide adaptive behaviour (Box 1). Throughout development, people tune their learning computations to the structural regularities of their environments — a process referred to as 'learning to learn'[8].

Goal-directed behaviour depends on learning via trial and error to take actions to maximize reward, a process formalized within reinforcement learning models. Environmental structure shapes not just what people learn but also how people learn from the outcomes of their actions, influencing, for example, how they weigh positive versus negative events[9] or more recent versus more distant experiences when updating their beliefs[10–12]. Understanding how variation in learning environments shapes how people learn to learn from reinforcement might be central to understanding why individual differences in reward learning emerge across development. Recognizing the importance of this question, developmental work has investigated how exposure to early-life adversity influences both the development of reward-related neurocircuitry and reward-learning behaviours[13,14]. This research has shown that the real-world reward statistics that people experience over months and years early in life exert influences on reward-learning processes that persist into adulthood[15,16]. However, the varied findings from this research are difficult to unify — it is unclear how and why specific types of experiences influence reward learning across childhood, adolescence and adulthood.

To make progress on understanding how the structure of experienced environments shapes reward learning across development, we can turn to systematic, computational accounts of the effects of environmental statistics on reinforcement learning. Researchers have developed normative models that explain how specific reward statistics of the learning environment (such as reward volatility, reward controllability and reward rates) should influence the optimal 'settings' of learning parameters. In addition, empirical work has revealed that the behaviour of adults largely conforms to these idealized predictions[10,17]. Such models of adaptive reinforcement learning have focused on the flexible calibration of learning in response to information encountered within single tasks over short timescales of experience, like the minutes or hours it takes participants to complete a laboratory experiment. However, they largely do not explain how past experiences, accumulated across multiple, varied environments over longer timescales, give rise to systematic variation in learning performance.

To bridge the timescales of real-world experience with the precise and theoretically motivated accounts provided by models of adaptive reinforcement learning, we consider the development of reward learning computations through the lens of models of meta-learning — and models of meta-reinforcement learning in particular. Meta-reinforcement learning models typically learn from reward on two timescales, often referred to as 'inner' and 'outer' loops[18]. In the inner loop, the model learns to perform a single task, such as navigating to a specific goal state[19] or choosing between different options to earn the most reward[20]. In the outer loop, through training on multiple tasks over longer timescales, models learn to improve how they learn within the shorter timescales of the inner task loops[18,21–26]. Models can learn to learn through varied mechanisms[18,25,27,28], including learning how to initialize[29] and update[19,30] the parameters that govern the inner loops of learning or discovering new inner-loop learning algorithms altogether[20,22]. Through training on multiple tasks, meta-reinforcement learning models gain the ability to solve related learning problems quickly, without extensive experience with the specific task at hand[25,27,31]. Akin to how humans and other animals learn to learn, meta-learning models learn at multiple levels of abstraction — they extract and leverage broader features of their training environments to specialize how they learn, while also retaining the flexibility to respond to the unique aspects of each individual task that they encounter[28,31].

In this Perspective, we argue that meta-learning provides a fruitful theoretical framework for understanding how and why experience shapes development. We suggest that considering development through this lens can inspire new empirical and computational research directions to best elucidate how the structure of early-life environments influences developmental trajectories of reinforcement learning computations. We begin by reviewing findings that suggest that individuals learn about the structural regularities of their environments across development and then use those regularities to guide how they learn to solve new problems. We then focus on computational modelling work that has demonstrated that specific statistical features of the reward environment influence adaptive reinforcement-learning computations. Next, we turn to developmental research examining how early-life experiences shape how people learn in diverse environments. We highlight the challenges of bridging theoretical computational work with studies of real-world environmental influences on development. We argue that conceptualizing development through the lens of meta-learning might promote productive cross-talk between these disparate fields, leading to a better understanding of how different aspects of experience influence how people learn to learn across the lifespan. Finally, we provide recommendations for how future work can best support the iterative construction of precise, computational accounts of how early-life experience shapes the development of adaptive behaviour.

Adapting learning to task statistics

Myriad laboratory studies have revealed that people learn to learn while completing experimental tasks: they exploit task structure to accelerate their acquisition of effective problem-solving and learning strategies. Learning to learn emerges early in development and persists across the lifespan. Across different domains, infants, children and adults can simultaneously solve specific learning problems while extracting higher-level representations of problem spaces that enable them to more effectively respond to new learning challenges[8,32–39]. For example, infants can extract and apply hierarchical rule structure to make correct inferences when faced with novel problems[38,40]. Children can discover and apply effective strategies for causal hypothesis testing[41,42], and they can learn the general features of a reward learning task, which in turn accelerates their acquisition of novel action–reward contingencies[8]. Critically, the derivation and application of learned strategies to guide subsequent learning is effective because the problems that people repeatedly encounter share similar structural features. Thus, although different problems require different solutions, learners can learn how to learn by exploiting shared structure across them.

Studies that have used computational models to characterize how people learn from reinforcement have revealed that specific statistical

Box 1

Neuroconstructivism

Our proposal builds on neoconstructivist and, in particular, neuroconstructivist frameworks that suggest that the developing brain continuously adapts to the structure of experienced environments[184–187]. Neuroconstructivism posits that the brain is evolutionarily endowed or genetically 'preprogrammed' with particular constraints. However, rather than precisely dictating the developmental trajectory of the organism, these constraints interact with each other and the experienced environment to shape neural activity, the wiring of neural networks and, ultimately, the nature of the mental representations that underlie learning and adaptive behaviour[186–188]. A central tenet of this framework is progressive specialization — the idea that new mental representations emerge as an individual encounters particular learning challenges within their current context. Thus, there is no top-down, adult-like goal state toward which a child develops; instead, changes in how individuals think and learn reflect adaptations to their immediate physical and social learning environments[186].

Many researchers have suggested that neural network models can provide insight into the algorithms that underlie the developmental learning processes proposed by neuroconstructivism[187,189–192]. The architecture of a neural network, including the number of artificial neurons and the existence of connections between them, reflect innate 'constraints' on development, whereas experience-induced changes in the strength of the connection weights reflect gradual adaptation to the learning environment[186,193]. Constructivist algorithms can further simulate how experience might change neural architecture itself[192]. For example, in one algorithm, a network first learns to maximize performance on a task with a fixed architecture and then recruits additional neural units or layers to further minimize errors[194].

We suggest that meta-learning architectures and algorithms can similarly provide insight into how experiences that unfold over developmental time shape the nature of the neurocognitive representations that underlie goal-directed behaviour. By explicitly considering how learning on multiple timescales is shaped by specific reward statistics, these models provide accounts of how and why different aspects of experience lead to lasting influences on reinforcement-learning computations.

features of the reward structure of an environment systematically shape how people learn to learn. In general, people can learn to take actions that bring about beneficial outcomes by increasing (or decreasing) their estimates of the value of different actions. They do so on the basis of the outcome that each action elicits: if an action elicits more reward than expected, then the estimated value of the action should increase, whereas if it elicits less reward than expected, then the estimated value of the action should decrease[43,44]. Although this general learning strategy can promote the selection of adaptive actions across contexts, differences in the precise dynamics of how value estimates are updated and used to guide future choices can dramatically alter patterns of action selection in ways that can be more or less beneficial across different contexts.

As with other types of learning problem, learning to reinforcement learn involves exploiting the environmental statistics of distinct contexts to optimize subsequent learning strategies. The optimization of learning involves adjusting the algorithms and the 'settings' of the parameters that govern reinforcement learning[45,46]. Theoretical and computational work on learning to reinforcement learn – which we refer to as 'adaptive reinforcement learning' – has demonstrated that the precise ways in which environmental structure tunes learning computations largely align with normative accounts of how idealized learners should tailor learning to gain the most reward across varied contexts[9,10,17,47].

Here, we focus on environmental statistics that might be particularly variable across the contexts that infants and children experience (Table 1). In particular, we consider the volatility of action–outcome contingencies, the overall reward rate or prevalence of positive versus negative rewards in the environment, and the controllability of reward outcomes. For each statistic, we first discuss theoretical work characterizing how the environment should normatively influence optimal learning algorithms and then examine empirical research that has tested these predictions within experimental tasks. We primarily review data from adults, with whom the majority of research on adaptive reinforcement-learning computations has been conducted, but we highlight data from children and adolescents where it exists.

Adapting learning to reward volatility

In many learning contexts, the relations between actions and the outcomes they elicit are volatile, such that they change dynamically over time. For example, an infant might learn that their cries sometimes elicit caregiver attention but are at other times ignored[48]. In volatile environments, the most recently experienced outcomes are the best indicators of the value of taking a particular action, because they best reflect the current state of the environment. Thus, learning to make good choices in volatile environments involves heavily weighing recent outcomes, effectively overwriting the influence of more temporally distant experiences. However, in environments with more stable reward contingencies, it tends to be computationally optimal for people to weigh recent outcomes less heavily, and instead to integrate over a longer history of experienced outcomes when determining the value of different actions – otherwise recent outcomes that arise owing to other sources of variability in the environment (such as stochasticity or noise) might cause them to suboptimally value recent actions[49–54].

Learning to learn involves estimating the underlying volatility of the environment from experience and using those volatility estimates to modulate the extent to which recent outcomes are weighed during learning (the learning rate). There are multiple computational accounts of how an idealized learner should solve this challenge[10,55–60]. In general, the learner estimates their own uncertainty about the value of their actions as well as the source of their uncertainty (for instance, the volatility of the environment) and increases their learning rate when they

Table 1 | Effects of environmental statistics on adaptive reinforcement learning

Environmental statistic	Predicted influence on reinforcement learning	Empirical findings	
		Behaviour	Neural mechanisms and physiological signatures
Volatility	Learning rates should be higher in environments with high volatility of action–outcome contingencies and lower in environments with low volatility of action–outcome contingencies[49–54]	Adults demonstrate higher learning rates when the environment is more volatile and lower learning rates when the environment is less volatile (in periods of stability)[10,12,55,56]	Neurons in the locus coeruleus respond to unexpected uncertainty triggered by context changes, releasing norepinephrine[179–181] and inducing network plasticity[182] and adjustments of learning rates[183]; pupil dilation in adults increases following changes in the reward contingencies of an environment, reflecting neural activation within the locus coeruleus[12,61]
Reward rates	Positive learning rates should be higher when environmental reward rates are low and negative learning rates should be higher when environmental reward rates are high[62]	Unclear; one study did not find evidence for learning rate adaptivity in adults[65]	Dopaminergic[64,67,75] and serotonergic[76–79] systems might influence how people learn from reward versus punishment, although the precise mechanisms remain unclear
Controllability	Learning rates and reliance on instrumental learning systems[17] should be higher in more controllable environments and lower in less controllable environments	Adults adjust the extent to which they update their beliefs about the value of their actions[17,84] on the basis of controllability estimates; adolescents[86] and adults[85–88] adjust valenced learning rates on the basis of the differential controllability of positive and negative outcomes	Estimating control might rely on computations implemented in the striatum and medial prefrontal cortex[83]; use of control estimates to modulate learning behaviour is likely to engage prefrontal cortical circuitry[88,90]

attribute their uncertainty to environmental volatility. Multiple studies have found that adults tend to adjust their learning rates as predicted by these optimal models, demonstrating higher learning rates when the environment is more volatile and lower learning rates during periods of stability[10,12,55,56]. For example, in one laboratory experiment, adults showed lower learning rates when repeatedly choosing between two options with reward probabilities that remained stable for 120 trials and higher learning rates when the probabilities reversed every 20 trials[10]. Further, adults also show physiological and neural signatures of tracking environmental volatility[12,61] (Table 1).

Theoretical and experimental work on learning in volatile environments suggests that, like optimal Bayesians, adult learners can track the uncertainty in their beliefs about the outcomes of different actions and the underlying environmental volatility that might give rise to that uncertainty. Learners can then effectively use volatility estimates to dynamically set the parameters that control their learning process over the minutes of experimental tasks.

Adapting learning to reward rates

Throughout their lives, learners face some environments that are rich in potential rewards and others with sparser opportunities to experience good outcomes. Computational theories suggest that learners can efficiently select the best actions across these varied environments by adapting how they weigh positive and negative feedback on the basis of the overall prevalence of reward[62]. Efficient learning of optimal action selection can be achieved through the misrepresentation or distortion of true action values[62,63]. Specifically, updating beliefs to different extents following negatively versus positively surprising events – having asymmetrical negative and positive learning rates, which we refer to as 'valenced' learning rates – distorts value estimates either downwards or upwards, depending on the direction of the asymmetry[64]. When the reward rate of an environment is high and many actions are frequently rewarded, then negative outcomes are a better signal to discriminate between similarly valued options[62]. As such, a pessimistic agent, who places more weight on negatively surprising versus positively surprising outcomes, will learn distorted value estimates that better differentiate high-value options and will make more optimal choices overall. The reverse is true when rewards are sparse – in that case, an optimistic agent will be better able to differentiate multiple low-value options and make more optimal choices[62].

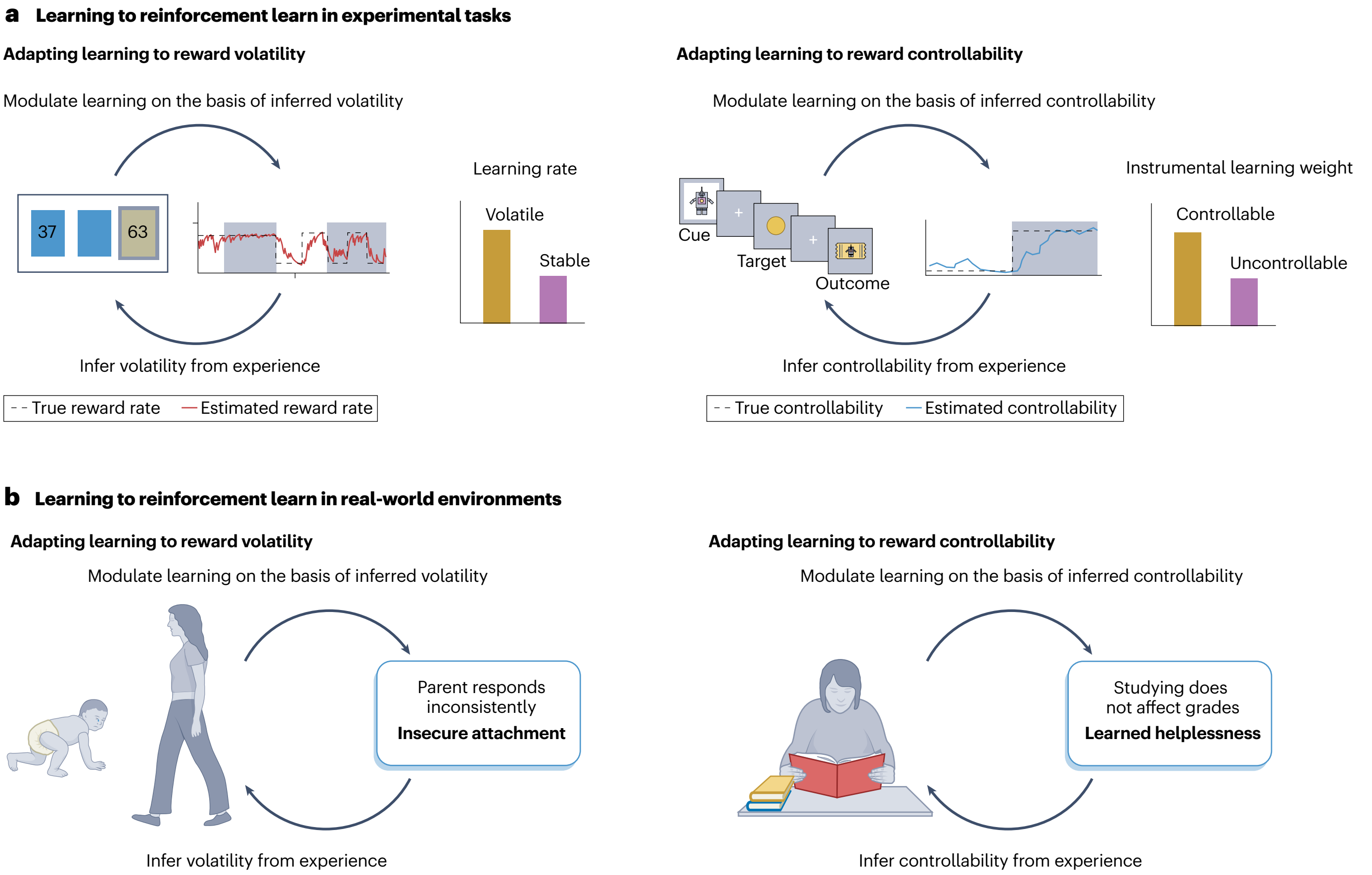
Across contexts, people face the challenge of learning how to learn from positive and negative outcomes to most effectively guide choice. Evidence for the ability of children, adolescents and adults to calibrate valenced learning rates to the reward statistics of an environment is mixed[9,65]. However, an extensive body of research suggests that people learn differently from positive and negative prediction errors[65–70], suggesting that they do tend to learn distorted value estimates across contexts. Further, these asymmetries vary across different learning environments – although the majority of learning tasks elicit positive learning rate asymmetries[68,71–74], others elicit negative learning rate asymmetries[65,69,70]. In one experiment, for example, children, adolescents and adults made repeated choices between four options that yielded different distributions of points across 2 blocks of 100 choices; participants demonstrated more positive learning rate asymmetries during the block in which distorting prediction errors upward led to greater reward gain, whereas this bias was attenuated in the block in which distorting prediction errors downward yielded more reward[9]. Setting asymmetric learning rates might involve the engagement of different neuromodulatory systems, including those that control dopamine[64,67,75] and serotonin[76–79]. Future work is needed to clarify the precise neuromodulatory mechanisms through which environmental statistics relate to valenced learning rates.

Adapting learning to reward controllability

Optimal learning strategies also differ depending on the degree of control an individual has over reward outcomes. Broadly, an individual should devote energy to demanding learning computations only in situations in which their actions can influence the outcomes they experience (that is, in which their actions are instrumental)[80].

Further, an individual should update beliefs about action values to a greater extent following outcomes that were likely to have been caused by those actions. An ideal learner can estimate the controllability of reward outcomes by comparing the probability of experiencing an outcome given that they took a specific action with the probability of experiencing an outcome given that they did not take that action[81]. If these probabilities are equal for all actions, then the outcome is not controllable. As these probabilities diverge, an agent can infer greater control and should increase the extent to which they learn about the value of their actions from experienced outcomes[17]. For example, if a student worked diligently on an essay and then received a good grade, they should increase their estimate of the value of hard work and continue to put forth effort on future assignments. However, if the student frequently receives poor grades after working hard, they might begin to believe that they have no control over their grades, and therefore stop adjusting the effort they put into their work.

Multiple laboratory studies have found that people estimate the degree of control they have over their environments[82,83] and modulate the extent to which they update their beliefs about the value of their actions accordingly[17,47,84] (Fig. 1). For example, in one experiment, children, adolescents and adults completed 180 trials of a reinforcement-learning task in two different controllability conditions[47]. In the controllable condition, participants could learn whether withholding or making a response to each of four different stimuli would lead to more reward gain; in the uncontrollable condition, reward outcomes were not contingent on participant responses. Computational modelling results revealed that participants across ages relied more on 'default' response strategies in the uncontrollable condition, whereas they relied more on instrumental learning of stimulus–action values when their actions were causally efficacious. Adults also adjust valenced learning rates on the basis of the differential controllability of positive and negative outcomes – when positive outcomes are more controllable than negative outcomes, adults demonstrate a positive learning rate asymmetry, updating their beliefs to a greater extent following positive rather than negative prediction errors. When negative outcomes are more controllable



a Learning to reinforcement learn in experimental tasks

Adapting learning to reward volatility

Adapting learning to reward controllability

b Learning to reinforcement learn in real-world environments

Adapting learning to reward volatility

Adapting learning to reward controllability

Modulate learning on the basis of inferred volatility
37
63
Learning rate
Volatile
Stable
Infer volatility from experience
- - True reward rate
— Estimated reward rate
Modulate learning on the basis of inferred controllability
Cue
Target
Outcome
Instrumental learning weight
Controllable
Uncontrollable
Infer controllability from experience
- - True controllability
— Estimated controllability
Modulate learning on the basis of inferred volatility
Parent responds inconsistently
Insecure attachment
Infer volatility from experience
Modulate learning on the basis of inferred controllability
Studying does not affect grades
Learned helplessness
Infer controllability from experience

Fig. 1 | Learning to reinforcement learn over multiple timescales. a, Within experimental tasks, participants adapt their learning computations to the reward statistics of the learning environment (including volatility[10] (left) and controllability[47] (right)), incrementally adjusting how they learn, enabling the more rapid acquisition of reward in accordance with the predictions of idealized Bayesian learning models. **b**, The same 'learning to learn' processes that have been shown to play out in experimental tasks might also influence learning across the lifespan. For example, an infant might experience inconsistent caregiver responses to their cries, leading them to infer that the environment is volatile (left). This experience might lead them to adjust how they learn from caregiver responses and lead to 'insecure attachment', characterized by long-lasting changes to how they learn from and form relationships with others. Similarly, a student who consistently receives poor grades on exams despite studying diligently might learn that their exam outcomes are not controllable (right). They might therefore reduce the extent to which they learn from exam feedback and eventually enter a state of 'learned helplessness' in which they stop trying to influence their grades at all. Part **a** adapted from ref. 10, Springer Nature Limited.

than positive outcomes, adults demonstrate a more negative learning rate asymmetry[85–88]. Estimating control might rely on computations implemented in the striatum[89] whereas the use of these control estimates to modulate subsequent learning is likely to engage prefrontal cortical circuitry[90,91] (Table 1).

Taken together, this body of research suggests that people learn how to reinforcement learn by using the statistics of their environments to flexibly modulate the computations that govern the learning process. The process of adaptive reinforcement learning described in the previous sections provides accounts of how different environmental statistics influence reward learning in distinct ways, but share central features: the learner simultaneously estimates key statistical properties of the environment while using those estimates to adjust learning parameters accordingly. These parameters determine how the agent learns which actions to take to gain reward. In this way, the beliefs of an agent about the structure of the environment ultimately influence the information that they subsequently encounter, which in turn iteratively shapes those beliefs. This work provides an important theoretical foundation for considering how these types of environmental statistics might influence learning over longer timescales, which we discuss in the remainder of the article.

Adapting learning to real-world statistics

Whereas the work reviewed in prior sections highlights how the learning of children, adolescents and adults aligns with the predictions of optimal models, in many laboratory experiments, the behaviour of people diverges from these normative predictions. For example, in some tasks, people demonstrate optimistic learning biases[92] even when such biases interfere with adaptive choice[9]. Similarly, people tend to demonstrate a belief that the outcomes in their environments are more controllable than they actually are[93]. Individuals with anxiety commonly overestimate environmental volatility and fail to calibrate their learning rates optimally to the underlying reward statistics of the environment[11,12]. At first glance, the persistence of maladaptive optimism and heightened controllability or volatility beliefs seem to be failures of learning to reinforcement learn. However, these 'failures' can be understood as learning to learn across longer timescales. Persistent biases in the settings of reinforcement-learning parameters, such as maladaptive optimism and heightened controllability biases, might be implemented by learning systems that have adapted to environments in which optimism and control beliefs are generally beneficial[94–96]. If these biases result from learning that takes place across many environments over many years, then the learner might not be able to overcome them within the short timeframe of a task. Instead, the architecture of the learning system might limit how effectively its dynamics can adapt to local environmental statistics. Thus, models of learning to learn across developmental timescales must be able to account not just for how an optimal learner can solve a specific learning task but also for how systematic deviations from optimality can get 'baked into' the learning system through experience — and, critically, why early experiences might be particularly influential.

In the next sections, we discuss how learning to learn may proceed within real-world environments over developmental timescales. We first highlight mechanisms of plasticity in the developing brain that enable early experiences to lead to long-lasting consequences for neurocognitive development. We then turn to research on early-life adversity and examine how early experiences with different reward statistics may influence reward-learning processes across childhood, adolescence and adulthood.

Plasticity in the developing brain

Extensive evidence suggests that the brain becomes increasingly specialized to process and learn from the specific input encountered during developmental sensitive periods[97,98]. During these periods, the brain is particularly malleable and responsive to environmental experience[98–100]. Sensitive periods enable the tuning of neurobiological circuitry to represent and process the idiosyncratic statistics of the environment of an organism[99,101,102]. This specialization process, and the eventual stabilization of representations, has benefits for later behaviour[103]. For example, infants who showed stronger neural discrimination of phonemes from their native language at 7 months of age demonstrated greater language abilities 2 years later, suggesting that early neural adaptation to the statistics of the environment tunes neurocognitive processing to accelerate learning[104]. Thus, sensitive periods can be thought of as learning to learn — the brain tailors its computations to learn more effectively in expected future environments.

Potential sensitive periods for reward learning[14,105], in which the reward statistics of early environments exert a pronounced influence on developmental outcomes, are less well-characterized than sensitive periods for sensory processing[3,106–109]. Although many studies have demonstrated that neural components of reward circuitry are sensitive to experience, it is less clear when and how experience with specific environmental statistics shapes the processing architecture of the brain in ways that might influence later reward learning.

Learning from early reward statistics

The same statistics that influence learning on short timescales (including volatility, reward rate and controllability) might also influence learning on longer timescales, altering developmental trajectories of reward-learning circuitry and instilling lasting biases that influence learning computations throughout the lifespan. We can look for evidence in support of this idea by examining research on how exposure to varied environments early in life leads to later consequences for learning to learn from reinforcement.

Most research on the influence of the early environment on reward learning focuses on the effects of adverse experiences in childhood (Box 2). Theoretical frameworks have posited that deprivation and threat constitute distinct subtypes of adversity[110]. Deprivation encompasses experiences of restricted environmental input, including reduced cognitive stimulation or social interaction, whereas threat involves either witnessing or directly experiencing harm or potential harm (such as physical violence or abuse)[111]. Experiences of deprivation and threat have been linked to changes in both reward circuitry and reward-guided behaviour[13,111].

Childhood experiences of neglect and abuse relate to blunted activation in key reward-processing regions, including the ventral tegmental area and striatum during reward anticipation and receipt, in adolescence and early adulthood[112–114]. Early experiences of deprivation and threat have also been associated with changes in patterns of connectivity between these subcortical structures and the prefrontal cortex that emerge in early childhood and persist through late adolescence[115–117]. Although deprivation and threat do not directly map to volatility, reward rates and controllability, this research suggests that the reward statistics of the early-life environment do indeed shape the neural architecture of reward learning systems.

The nature of the consequences of early adverse experiences for reinforcement learning and reward-guided behaviour more generally is not fully clear. For many years, a predominating perspective was that early-life adversity leads to 'deficits' in later learning behaviours[118].

Box 2

Learning across sociocultural contexts

Our Perspective focuses largely on how 'adverse' or stressful experiences early in life might shape reward learning computations. However, the sociocultural environments in which children grow up also shape the reward statistics that they experience early in life and similarly exert lasting influences on how they learn and make decisions[195,196]. Research has revealed that the choice strategies of children adapt to the statistics of their social and cultural contexts, such that they are modulated by the reliability of caregivers in their environments[197], normative cultural practices[198], and broader societal and economic forces[199]. For example, children from countries with higher market integration, which tends to yield higher food security, demonstrate a stronger willingness to wait for larger rewards versus accepting smaller rewards immediately as well as a greater propensity to take risks. Greater market integration might attenuate the potential harms of uncertainty[200], leading to differences in the reward landscape across societies that ultimately shape the learning and decision-making strategies of children.

This type of research highlights how the sociocultural contexts in which children develop shape learning to learn. Future work should continue to address how practices and norms that differ across cultures influence specific reward statistics of the environment and, in turn, reinforcement-learning computations. For example, it might be the case that some cultures emphasize the agency of children more than others[201], such that caregivers within certain societies offer children more opportunities to control the rewards they experience. This type of variation could be used to ask how opportunities for control shape biases toward instrumental learning in different environments across development. To date, the majority of research on the development of reward learning computations has focused on 'WEIRD' (Western, educated, industrialized, rich and democratic) populations, but a comprehensive theory of development as a process of learning to learn must account for variation in the development of learning and choice behaviours in diverse societies across the world[202,203]. This vision could be realized by drawing on methods and perspectives from sociology, anthropology, economics and the broader field of childhood studies.

However, newer work has highlighted the adaptive nature of these developmental responses, focusing on how adverse early-life environments promote learning behaviours that are well-suited to the challenges that these environments pose[119–124]. Apparent 'deficits' or maladaptive behaviours emerge when there is a mismatch between the early environments that shape learning architectures and the learning environments encountered later in development[125,126]. For example, children raised in institutional care settings demonstrate impulsive and exploitative decision strategies[127], which are disadvantageous in stable environments, but might yield more reward in uncertain contexts with statistics that more closely match those that they experienced early in life[123]. It might be the case that adaptation to the statistics of early-life environments leads to biases in the tuning of learning parameters that persist in contexts encountered later.

Given the heterogeneity in measures and tasks used to examine how early-life adversity influences reward learning, it is difficult to extract generalizable principles. Nevertheless, across multiple studies, experiences of early-life adversity have been shown to disrupt later reward learning, such that children, adolescents and adults who experienced greater adversity and more stress early in life demonstrate slower learning of rewarding actions relative to their peers[15,16,128–130]. For example, adolescents who had experienced physical abuse were worse than their peers at selecting rewarding actions in a probabilistic reward learning task – they represented the value of the better and worse options as closer together and also made more random choices[129]. Similarly, adolescents who had had high levels of stress in early childhood were worse than their peers at learning initial action–outcome contingencies as well as at updating their beliefs when contingencies changed[128].

Beyond the central finding of early-life adversity leading to disruptions in reward-guided behaviour, features of different early-life environments have been linked to varied aspects of reinforcement learning. For example, in one study, normative variation in early-life adversity led to increases in positive learning rates: children aged 9–12 years with more frequent and intense experiences of adverse life events increased their value estimates to a greater extent following recent, positive prediction errors relative to children who had fewer adverse experiences[131]. In another study, children aged 8 and 9 years with greater exposure to stressful events were more likely to avoid cues that elicited both positive and negative outcomes but only if they had experienced high levels of perceived social isolation[132]. However, in other research, children and adolescents aged 10–15 years who had experienced maltreatment demonstrated similar behaviour on a reward learning task to children who had not, although their learning-related patterns of neural activation in both subcortical reward structures and frontal cortex differed[133].

Studies of the effects of early-life adversity on neurocognitive development suggest that the reward statistics of early-life experiences clearly shape the architecture of reward learning systems and have functional consequences for adaptive behaviour. However, current computational accounts cannot explain why individual differences in reinforcement learning emerge through interactions with different environments over development.

The lens of meta-reinforcement learning

Learning to learn from the statistics of experience could enable biological and artificial agents to strike a balance between efficiency and flexibility. Evolutionarily hard-wired 'solutions' to challenges posed by consistent features of the environment enable agents to most efficiently orchestrate their behaviour[134]. However, these innate solutions limit the flexibility of agents to adapt to the specific environmental demands that they face. Learning how to behave from experience enables flexibility but learning from the outcomes of actions can be slow – artificial agents often require extensive training to be able to learn effective policies for action selection[135]. Meta-learning architectures balance efficiency and flexibility by enabling learning systems to gradually specialize to the types of environments they have encountered, leading to increasingly adaptive learning when the future reflects the past[18,27]. Considering development through the lens of meta-reinforcement learning therefore bridges the precise computational accounts of adaptive reinforcement learning with the real-world, developmental timescales of the early-life adversity literature.

The developmental process can be conceptualized through this meta-reinforcement learning lens (Fig. 2). The environments that people encounter early in life can be thought of as the distribution of environments on which a model is trained. Through an outer loop of learning, the brain slowly adjusts its architecture to most effectively respond to the environmental statistics it encounters, encoding learning biases or strategies that control the inner-loop learning algorithms that are leveraged to solve specific tasks. Over time, dynamic interactions with the environment shape the architecture of the brain, which in turn influences how organisms respond to new learning challenges. By conceptualizing the development of reward-learning computations through the lens of meta-reinforcement learning, existing models can be modified to test specific theoretical predictions about how variation in the statistical structures of early-life environments might confer learning biases that persist across the lifespan. Considering early-life environments as 'tasks' that comprise the training sets for models of meta-reinforcement learning opens several interesting avenues for further investigation. For example, researchers could train models on different distributions of learning tasks and examine whether they exhibit particular biases in their parameters when tested on new learning problems[136]. By training models for different durations of time, researchers could attempt to recapitulate developmental variance in both overall performance at testing as well as in specific patterns of learning errors. Finally, researchers could probe whether the representational content of the network at different points in its training regimen recapitulates patterns of developmental change in neurocognitive representations[137]. To begin to hone these research directions and translate meta-reinforcement learning from a conceptual framework into an empirical research agenda, researchers must consider which model architectures, algorithms and training data best reflect what is known about the developing brain and the environments in which it learns.

Developing model architectures and algorithms

Meta-learning models learn to reinforcement learn through different mechanisms, which provide testbeds for different developmental theories (Box 3). Meta-learning models are frequently instantiated as recurrent neural networks, in which layers of artificial neurons pass information both forward to subsequent layers of the network and back onto themselves, enabling the network to 'remember' and learn from sequential input[138]. There are many approaches to implementing meta-learning in recurrent neural networks[20,22,29,139–145]. Through gradual training over multiple tasks, a network can learn how to optimally initialize its parameters (the connection weights between artificial neurons and the biases or 'offsets' that influence activation within each neuron independent of their input) to rapidly reach effective settings for new tasks that it encounters[29]. Other meta-learning algorithms also enable the network to learn how much and in what direction to update its parameters after each batch of training[19] – to learn meta-parameters that determine how the network learns from experience within individual tasks. Additionally, other networks learn reinforcement-learning algorithms that are implemented through recurrent activation dynamics[20,22]. Here, we highlight two different models of meta-reinforcement learning and explore how they could be modified to begin to address developmental questions.

In one biologically inspired recurrent neural network model, the 'prefrontal cortex' of the network dynamically adjusts its organization via the learning of connection weights that are updated slowly on the basis of experienced rewards – in other words, it learns in an outer loop via a mechanism akin to synaptic plasticity[20]. At test, when the network is exposed to an experimental task, its learned weights remain fixed but it can flexibly adjust how it learns via recurrent activation dynamics,

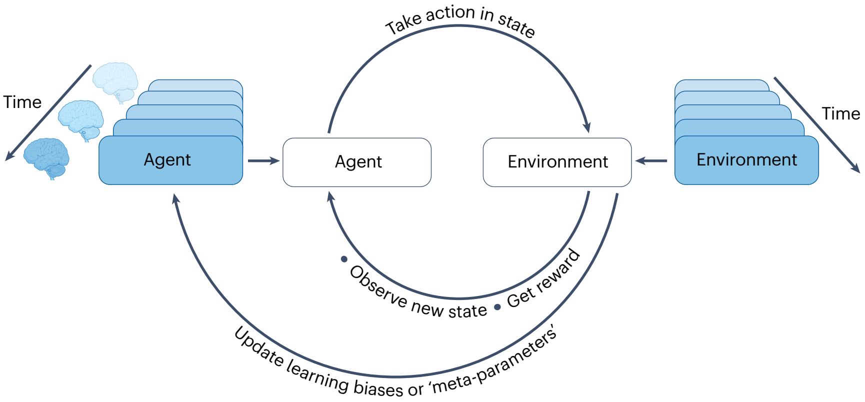


Fig. 2 | Development through the lens of meta-reinforcement learning. The developmental process can be conceptualized through the lens of meta-reinforcement learning. Across development, learning from the outcomes of actions proceeds across multiple timescales. Through 'inner loops' of learning, agents learn to tailor their actions to the demands of specific environments on relatively short timescales. Within these inner loops, agents take actions in specific states, observe the outcomes of those actions and update the beliefs that guide their future decisions. After observing the outcome of an action in a particular state, an agent also updates how they learn through an 'outer loop' of learning. Through this outer loop, the agent slowly adjusts how they learn to respond more effectively to the statistics of the environment encountered. These inner-loop and outer-loop learning processes proceed across multiple environments over time, such that the learning biases that an agent encodes might be tuned over long timescales of experience. Importantly, different environments might be encountered at different points in development. In this way, 'learning to learn' mechanisms could shape the neurocognitive developmental trajectory of the agent.

Box 3

Modelling human-like cognitive development

Beyond learning over multiple timescales, there are many other ‘ingredients’ that researchers have integrated into artificial agents to make their learning and behaviour more closely resemble the learning that proceeds over human development.

Inductive biases

Human learners might have evolved preprogrammed ‘inductive biases’ that influence information processing from the earliest days of life[204,205]. Incorporating these biases into machines might accelerate their learning and make it more human-like[206]. However, the extent to which knowledge is innate, and therefore should be built into human-like artificial agents, is an area of active debate[206]. It is unclear whether early-emerging biases in reinforcement learning (such as an ‘optimism’ bias[94]) are learned through experience or inherited through evolutionary mechanisms. Although the focus of our Perspective is on learning mechanisms that operate across development, meta-learning also proceeds on evolutionary timescales[27,207]. Constructing a mechanistic account of the role of experience in shaping reward learning computations requires understanding what aspects of behaviour might not be learned within the lifetime of an individual.

Changes to network architectures

The human brain changes dramatically over the first two decades of life[208]. Understanding this biological development and incorporating parallel changes into neural network architectures during extended training periods might yield more robust learning and performance that more closely resembles human learning[192]. For example, one study incorporated the heightened neurobiological pruning of excitatory synapses that occurs during adolescence into neural network models while they were being trained to solve reinforcement-learning and working memory tasks[209]. Doing so improved aspects of the performance of the networks that paralleled the developmental changes typically observed in studies of human behaviour. Other work has found that inducing computational ‘noise’ in recurrent neural networks improves their performance[210,211], particularly in exploring and generalizing to novel contexts[212–214]. The ‘noisiness’ or variability with which the brain processes information might change across development[215] — incorporating similar changes in the ‘noisiness’ of networks across their training and test periods could potentially recapitulate performance improvements observed across development.

Curriculum learning and embodiment

When learning about their environments, humans actively shape the input they receive, creating a structured curriculum that benefits knowledge acquisition[216]. The natural ‘curriculum’ of human learners is inherently constrained by their sensory and motor development — infants gradually learn from more complex information as their visual acuity increases and they learn to crawl and walk, which leads to a dramatic expansion of the information available for them to process[216]. In addition, human learners use uncertainty to guide information-seeking and learning, preferentially attending to and exploring information that will accelerate their own learning progress[217,218]. Structured, non-random training curricula can also benefit the performance of machine learning models[219], and integrating different functional forms of ‘curiosity’ into embodied robots can similarly create self-organized, structured curricula that promote learning of complex behaviours[220,221]. Future research in both humans and artificial agents should investigate what a structured curriculum for reinforcement learning looks like and whether the development of more complex forms of reward-guided behaviour depends on sequential exposure to increasingly complex reward statistics.

harnessing ‘prefrontal cortex’ representations for trial-and-error, inner-loop task-learning. For example, when trained on environments with varying levels of volatility, the model learned weights that enabled it to dynamically adjust its learning rate even when, later, those weights remained fixed[20]. In this way, the network provides an account of how the prefrontal cortex might learn from reward via two distinct mechanisms (changes in synaptic weights and activation dynamics) that unfold over two distinct timescales, tying together empirical observations into a unified computational theory. Newer work has capitalized on the specific, testable predictions that emerge from this account, and provided evidence that prefrontal plasticity is essential for outer-loop learning to learn while prefrontal activation supports inner-loop trial-by-trial learning[146].

This model can potentially be modified to provide insight into central developmental questions. By systematically manipulating the architecture and training regimen of the model and validating model simulations with additional empirical experiments, researchers could gain insight into how experiences at specific developmental time points might alter prefrontal cortex architecture and lead to different biases in reinforcement learning behaviours at later time points. Biological systems are often characterized by differential plasticity within different circuits at different times[100], a feature that could be approximated by varying the rate at which model weights are updated over training episodes[147] and across specific connections. In addition, the opening and closing of sensitive periods in biological systems might be governed by experience – organisms with higher ‘uncertainty’ about the conditions they might encounter in the future can show more prolonged windows of plasticity[102,148]. To reflect this biological principle, the model could be modified such that the rate at which its connection weights change is not determined a priori by the experimenter but governed by some aspect of the representation of the environment and the uncertainty of the network in its selection of actions. Model simulations could reveal how these features interact with the statistics of the tasks on which the model is trained and provide insight into how early plasticity facilitates or constrains the later adaptability of learning systems to different environmental demands.

Other models that learn to reinforcement learn via different mechanisms could similarly be harnessed to ask developmental questions. Many model architectures characterize two timescales of learning via an outer-loop mechanism that extracts and exploits shared structural features across learning environments and a distinct inner-loop mechanism that learns how to select rewarding actions within individual tasks. However, the world is not neatly segmented into an outer loop of 'developmental time' and an inner loop of experimental tasks. Instead, individuals might parse their continuous stream of experience into contexts with a more complex hierarchy. For example, people might learn to meta-learn not only how to set the parameters that govern how they learn from reward within a specific context but also how to determine how to set those higher-level parameters. For example, the learning rate that determines how the weights of a network change with experience could itself be learned via a learning process that could be optimized for different environments. Model architectures and algorithms different from the specific recurrent neural network discussed above might be better suited to address how these many loops of learning interact. For instance, in meta-gradient reinforcement learning, models simultaneously optimize policies for selecting rewarding actions and the meta-parameters that control task-level learning[149–151]. Critically, this optimization of the meta-parameters is done via the same learning mechanism as the optimization of the inner-loop parameters: the meta-parameter (outer loop) updates can be mathematically expressed in terms of how they influence the inner-loop parameter updates. These meta-parameters can then be similarly optimized to enhance the performance of the model in accomplishing its objective[149–151]. This flexible mechanism to optimize meta-parameters could theoretically be applied recursively to explain learning across many different levels of abstraction.

By casting learning in this way, extensions of meta-gradient models could be exploited to ask about the nature of the many outer, inner and intermediate loops that might characterize learning at different time points in development. For example, across development, individuals might learn to learn at increasingly deep hierarchical levels, such that a model that learns to learn to learn best approximates the performance of older participants on a learning task, whereas a model with less flexibility is a better approximation of how children learn from reward. Alternatively, increasing experience might lead outer loops of learning to stabilize, such that the learning of children is best characterized by a model with a deeper hierarchy, leading to an outsized influence of early environments on later learning. These hypotheses must be constrained by additional data; however, they illustrate the types of developmental questions and research directions that meta-reinforcement learning models afford.

Developing model training regimens

Researchers can also sharpen theoretical predictions about how experience shapes the development of reward learning by varying the environments in which meta-learning models are trained. The environments that children experience early in life can profoundly influence later reward learning[13]. However, many questions remain about how the timing, duration and sequential order of experience with different reward statistics influence learning. To leverage meta-reinforcement learning models to gain insight into the developmental process, we must ensure that model training regimens recapitulate aspects of variance in real-world environments. Doing so will require both better measurement techniques and methods to translate the statistics derived from those measures of early-life experience into model input.

Studies examining how the early environment influences reward learning typically characterize early-life experiences via structured questionnaires that ask about the occurrence of particular types of events in their lives. For example, studies of the effects of early-life adversity on reinforcement learning have used measures, including the Early Life Stress Questionnaire[152] and the Youth Life Stress Interview[153], which ask participants about negative events they have experienced (such as physical abuse or domestic conflict) and associated stress. Survey and interview measures have been shown to relate to different aspects of behaviour across development[114–116]. However, these measures are difficult to relate to computational theories of reinforcement learning – and difficult to use to construct training environments for models – because the experiences they assess do not map cleanly onto the same types of statistics that are manipulated in reward-learning tasks.

Contemporary theories of early-life adversity have proposed that 'environmental unpredictability' shapes neurocognitive development[111,154]. Unpredictability can be assessed by asking participants about experiences, including parental job changes, moving homes or neighbourhoods, or shifts in family economic conditions[111]. Environmental unpredictability might be more directly related than threat or deprivation to the types of statistics that have been manipulated in reward-learning experiments. However, unpredictable events can arise for different reasons – positive and negative reward contingencies in the environment might be stochastic, volatile or uncontrollable – and the inferences of people about the underlying causes of the 'unpredictability' they experience might shape how they learn from and about unpredictable reward outcomes. For example, if a parent sometimes comforts their crying infant and sometimes ignores them, the infant might infer that they are in a volatile environment, such that the contingency between their cries and the behaviour of their parent changes rapidly, and that they should rely only on their most recent experience to determine the causal effects of their own actions. Alternatively, they might infer that they are in an uncontrollable environment and learn that there is no causal link whatsoever between their actions and the behaviour of their parent. The inferences the infant makes might determine the beliefs they develop about the structure of the world and the types of environments for which they specialize their reward-learning computations.

More finely grained measures of early-life reward statistics could be developed by extracting features of interest from live or recorded naturalistic interactions between children and their environments, as is done to study language learning[155], visual perception[156] and socioemotional development[157]. For example, researchers often observe interactions between children and their parents to examine parental rates of contingent responding[158,159]. Given the central role of parents in shaping the early learning environments of children, contingent responding could provide a proxy for environmental control. As children enter adolescence, parental influence on their learning environments diminishes[160] and therefore control over reward contingencies could be measured via experience-sampling methods[161] that probe the number and frequency of autonomous choices made throughout the day. Similar approaches could be taken to examine other reward statistics such as reward rates and the volatility of reward outcomes. For instance, researchers could code the valence of parental responses to children to estimate the relative proportion of positive versus negative feedback that children experience.

Beyond recorded video data, other sources of real-time information about the experiences of people (for example, GPS tracking[162,163]) could be used to measure the regularity of daily and weekly routines. Although variability in locations visited is not directly analogous to the volatility of the reward contingencies, these measures might provide insight into the extent of variation in reward landscapes that individuals encounter, which could inform the design of different distributions of training tasks that meta-reinforcement learning models perform.

These natural statistics could then be mirrored within simple reward-learning tasks that models could complete. By varying the distributions of tasks on which meta-reinforcement learning models are trained, researchers could simulate how different types of experience influence the ability of a learning system to adapt to the environments it later encounters. For example, researchers could test whether existing meta-learning models can adapt to both volatile and stable environments when trained mostly on volatile environments. Importantly, the influence of training experience is likely to depend on the plasticity of the outer-loop learning parameters, which could be modified to recapitulate aspects of biological development. In this way, model simulations could help to untangle the complex relations between neural plasticity, experience in different types of environments and the ability of the learning system to adapt to diverse environments.

Conclusion

Understanding how environments shape how people learn is central to understanding how and why individual differences in goal-directed behaviour emerge. Models of adaptive reinforcement learning offer precise, theoretical accounts of how the reward statistics of the environment modulate learning computations. Developmental research on the influence of early adversity on later reward learning has suggested that variation in the reward statistics experienced early in life can lead to lasting changes in learning architectures. We suggest that models of meta-reinforcement learning can provide more incisive accounts of how the environment shapes goal-directed behaviour over time. These models couple the precise predictions of simpler models of adaptive reinforcement learning with the nested timescales that characterize cognitive development. By using these models to inspire new hypotheses and methods for empirical research and by using findings from empirical studies to iteratively refine model architectures, algorithms, and training data, researchers can better approximate the processes of developmental change in learning.

The iterative construction of models that recapitulate signatures of learning across development – and signatures of development itself – will also require well-designed animal studies in which key environmental statistics are manipulated, diagnostic tasks that can isolate learning from learning to learn, and broader consideration of the sociocultural contexts in which all human learning occurs. Reward statistics of early learning environments can be directly manipulated in animal studies, enabling researchers to alter specific properties of the environment and break the natural correlations that exist between reward statistics (such as volatility and reward rate) in most real-world contexts. Many studies have manipulated the rearing environments of rodents to create specific types of 'adverse' conditions. For example, studies have induced unpredictability in maternal caregiving behaviour[164–166], exposed animals to uncontrollable stressors or reinforcement[167,168], and manipulated the prevalence of cognitively enriching stimuli in animal environments[169,170]. Variations in animal-rearing environments have been linked to differences in the development of adaptive behaviour, including differences in memory performance[171], attachment learning[172] and repetitive, stereotyped responses[169], highlighting the insights that can be gained from controlled manipulation of the experiences of animals. Moving forward, models of meta-reinforcement learning could be used to generate hypotheses about how exposure to specific distributions of reward statistics shapes behaviour over time. These specific distributions could potentially be instantiated in the early environments of animals to determine their effects on learning across development.

In research with humans, developmental scientists can more rigorously test the specific predictions that emerge from model simulations by designing tasks that better capture how both outer and inner loops of learning change with age and experience. Infants, children, adolescents and adults extract and use structural regularities from their past learning experiences to guide how they approach new problems[27,173]. However, it is often difficult to isolate developmental changes in learning to learn (setting learning parameters) from developmental changes in learning (learning the optimal responses to make). Future studies can better isolate learning from learning to learn by having participants complete many sessions of similar learning tasks. Across sessions, tasks should share structural features that dictate how one should learn to learn but differ in the specific action selection policies; improvements in performance within a single session would provide an index of learning, whereas improvements across sessions would index learning to learn[8,146]. Importantly, the tasks used for such experiments would need to have adequate complexity such that the learning to learn challenge would require prolonged experience. Many of the tasks that have been used in existing studies were designed to illustrate how people learn to learn with only limited exposure to a particular environment, and might not effectively capture individual differences in the incremental adjustments of learning algorithms or parameters that proceed over longer timescales. In multi-session experiments with complex tasks, researchers could attempt to recapitulate the slow, experience-driven learning that proceeds over developmental timescales by varying the order and duration of the task environments to which people are exposed and measuring how such exposure instantiates different learning biases. By including participants across a wide, cross-sectional age range in studies with this design, researchers could examine differences in learning to learn across the lifespan.

Finally, although we have focused on experiments and models in which a single agent interacts with an environment devoid of other agents, the learning of children is embedded within social contexts. Simple reward-learning tasks do not capture the complexity of real-world experience, in which reward statistics are embedded in dynamic, multisensory and highly social contexts. The reward statistics that define the environments of children are largely determined by the behaviour of other people, including their caregivers and peers. The ultimate goal for developmentalists is to understand not only how specific reward statistics shape learning but also how the developing brain derives reward statistics from a rich tapestry of experiences. Addressing this challenge might require thinking creatively about how to use more naturalistic data sets to explore the nature of reward feedback itself across development by, for example, amending neural network models of reward learning with modules that translate emotional expressions from videos of caregivers or spoken words and phrases into teaching signals that can train the rest of the network[174–176]. By incrementally closing the gap between the input that children learn from in the real world and the environments in which models are trained, researchers can better probe the effects of natural variation in experience on learning to learn.

Finally, children are also taught to learn – a central goal of many formal educational systems is not simply to convey knowledge but to teach children to become expert learners[177,178]. Even adults surround themselves with cultural products such as books, videos and podcasts that teach them how to effectively learn across diverse environments. Ultimately, a complete characterization of learning to learn across development must address how bottom-up discovery of learning strategies interacts with top-down, explicit instruction. Expanding models of meta-learning to account for interactions between multiple agents can advance the understanding of how the sociocultural contexts in which learning systems are embedded are themselves shaped by learning on multiple timescales.

Published online: 18 April 2024

References

1. Scott, L. S., Pascalis, O. & Nelson, C. A. A domain-general theory of the development of perceptual discrimination. *Curr. Dir. Psychol. Sci.* **16**, 197–201 (2007).
2. Scott, L. S. & Monesson, A. The origin of biases in face perception. *Psychol. Sci.* **20**, 676–680 (2009).
3. Werker, J. F. & Tees, R. C. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant. Behav. Dev.* **7**, 49–63 (1984).
4. Hospodar, C. M., Hoch, J. E., Lee, D. K., Shrout, P. E. & Adolph, K. E. Practice and proficiency: factors that facilitate infant walking skill. *Dev. Psychobiol.* **63**, e22187 (2021).
5. Saccani, R., Valentini, N. C., Pereira, K. R., Müller, A. B. & Gabbard, C. Associations of biological factors and affordances in the home with infant motor development. *Pediatr. Int.* **55**, 197–203 (2013).
6. Sheridan, M. A., Peverill, M., Finn, A. S. & McLaughlin, K. A. Dimensions of childhood adversity have distinct associations with neural systems underlying executive functioning. *Dev. Psychopathol.* **29**, 1777–1794 (2017).
7. Amso, D., Salhi, C. & Badre, D. The relationship between cognitive enrichment and cognitive control: a systematic investigation of environmental influences on development through socioeconomic status. *Dev. Psychobiol.* **61**, 159–178 (2019).
8. Harlow, H. F. The formation of learning sets. *Psychol. Rev.* **56**, 51–65 (1949).
9. Nussenbaum, K., Velez, J. A., Washington, B. T., Hamling, H. E. & Hartley, C. A. Flexibility in valenced reinforcement learning computations across development. *Child Dev.* **93**, 1601–1615 (2022).
10. Behrens, T. E. J., Woolrich, M. W., Walton, M. E. & Rushworth, M. F. S. Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007).
11. Gagne, C., Zika, O., Dayan, P. & Bishop, S. J. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife* **9**, e61387 (2020).
12. Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X. & Bishop, S. J. Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat. Neurosci.* **18**, 590–596 (2015).
13. Hanson, J. L., Williams, A. V., Bangasser, D. A. & Peña, C. J. Impact of early life stress on reward circuit function and regulation. *Front. Psychiatry* **12**, 744690 (2021).
14. Galván, A. Neural plasticity of development and learning. *Hum. Brain Mapp.* **31**, 879–890 (2010).
15. Wilkinson, M. P., Slaney, C. L., Mellor, J. R. & Robinson, E. S. J. Investigation of reward learning and feedback sensitivity in non-clinical participants with a history of early life stress. *PLoS One* **16**, e0260444 (2021).
16. Birn, R. M., Roeber, B. J. & Pollak, S. D. Early childhood stress exposure, reward pathways, and adult decision making. *Proc. Natl Acad. Sci. USA* **114**, 13549–13554 (2017).
17. Dorfman, H. M. & Gershman, S. J. Controllability governs the balance between Pavlovian and instrumental action selection. *Nat. Commun.* **10**, 5826 (2019).
18. Botvinick, M. et al. Reinforcement learning, fast and slow. *Trends Cogn. Sci.* **23**, 408–422 (2019).
19. Li, Z., Zhou, F., Chen, F. & Li, H. Meta-SGD: learning to learn quickly for few-shot learning. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1707.09835 (2017).
20. Wang, J. X. et al. Prefrontal cortex as a meta-reinforcement learning system. *Nat. Neurosci.* **21**, 860–868 (2018).
21. Wang, J. X. et al. Learning to reinforcement learn. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1611.05763 (2016).
22. Duan, Y. et al. RL2: fast reinforcement learning via slow reinforcement learning. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1611.02779 (2016).
23. Weng, L. *Meta Reinforcement Learning* https://lilianweng.github.io/posts/2019-06-23-meta-rl/ (2019).
24. Langdon, A. et al. Meta-learning, social cognition and consciousness in brains and machines. *Neural Netw.* **145**, 80–89 (2022).
25. Binz, M. et al. Meta-learned models of cognition. *Behav. Brain Sci*. https://doi.org/10.1017/S0140525X23003266 (2023).
26. Schaul, T. & Schmidhuber, J. Metalearning. *Scholarpedia J.* **5**, 4650 (2010).
27. Wang, J. X. Meta-learning in natural and artificial intelligence. *Curr. Opin. Behav. Sci.* **38**, 90–95 (2021).
28. Lansdell, B. J. & Kording, K. P. Towards learning-to-learn. *Curr. Opin. Behav. Sci.* **29**, 45–50 (2019).
29. Finn, C., Abbeel, P. & Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proc. 34th International Conference on Machine Learning* (eds Precup, D. & Teh, Y. W.) 70, 1126–1135 (PMLR, 2017).
30. Doya, K. Metalearning and neuromodulation. *Neural Netw.* **15**, 495–506 (2002).
31. Griffiths, T. L. et al. Doing more with less: meta-reasoning and meta-learning in humans and machines. *Curr. Opin. Behav. Sci.* **29**, 24–30 (2019).
32. Behrens, T. E. J. et al. What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
33. Crowley, K. & Siegler, R. S. Explanation and generalization in young children's strategy learning. *Child Dev.* **70**, 304–316 (1999).
34. Bielaczyc, K., Pirolli, P. L. & Brown, A. L. Training in self-explanation and self-regulation strategies: investigating the effects of knowledge acquisition activities on problem solving. *Cogn. Instr.* **13**, 221–252 (1995).
35. Bakst, L. & McGuire, J. T. Experience-driven recalibration of learning from surprising events. *Cognition* **232**, 105343 (2023).
36. Dubey, R., Grant, E., Luo, M., Narasimhan, K. & Griffiths, T. Connecting context-specific adaptation in humans to meta-learning. Preprint at https://doi.org/10.48550/arXiv.2011.13782 (2020).
37. Verbeke, P. & Verguts, T. Humans adaptively select different computational strategies in different learning environments. Preprint at *bioRxiv* https://doi.org/10.1101/2023.01.27.525944 (2023).
38. Werchan, D. M., Collins, A. G. E., Frank, M. J. & Amso, D. 8-month-old infants spontaneously learn and generalize hierarchical rules. *Psychol. Sci.* **26**, 805–815 (2015).
39. Mark, S., Moran, R., Parr, T., Kennerley, S. W. & Behrens, T. E. J. Transferring structural knowledge across cognitive maps in humans and models. *Nat. Commun.* **11**, 4783 (2020).
40. Brown, A., Kane, M. J. & Echols, C. H. Young children's mental models determine analogical transfer across problems with a common goal structure. *Cogn. Dev.* **1**, 103–121 (1986).
41. Nussenbaum, K. et al. Causal information-seeking strategies change across childhood and adolescence. *Cognit. Sci.* **44**, e12888 (2020).
42. Kuhn, D. & Phelps, E. The development of problem-solving strategies. *Adv. Child Dev. Behav.* **17**, 1–44 (1982).
43. Rescorla, R. A. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning Curr. Res. Theory* **2**, 64–69 (1972).
44. Sutton, R. S. & Barto, A. G. *Reinforcement Learning. An Introduction* (MIT Press, 1998).
45. Kool, W., Gershman, S. J. & Cushman, F. A. Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychol. Sci.* **28**, 1321–1333 (2017).
46. Ruel, A., Devine, S. & Eppinger, B. Resource-rational approach to meta-control problems across the lifespan. *Wiley Interdiscip. Rev. Cogn. Sci.* **12**, e1556 (2021).
47. Raab, H. A., Goldway, N., Foord, C. & Hartley, C. A. Adolescents flexibly adapt action selection based on controllability inferences. *Learn. Mem.* **31**, a053901 (2024).
48. Salter Ainsworth, M. D. The Bowlby-Ainsworth attachment theory. *Behav. Brain Sci.* **1**, 436–438 (1978).
49. Diederen, K. M. J. & Schultz, W. Scaling prediction errors to reward variability benefits error-driven learning in humans. *J. Neurophysiol.* **114**, 1628–1640 (2015).
50. Payzan-LeNestour, E. & Bossaerts, P. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. *PLoS Comput. Biol.* **7**, e1001048 (2011).
51. Piray, P. & Daw, N. D. A model for learning based on the joint estimation of stochasticity and volatility. *Nat. Commun.* **12**, 6587 (2021).
52. Dayan, P., Kakade, S. & Montague, P. R. Learning and selective attention. *Nat. Neurosci.* **3**, 1218–1223 (2000).
53. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
54. Soltani, A. & Izquierdo, A. Adaptive learning under expected and unexpected uncertainty. *Nat. Rev. Neurosci.* **20**, 635–644 (2019).
55. Nassar, M. R., Wilson, R. C., Heasly, B. & Gold, J. I. An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *J. Neurosci.* **30**, 12366–12378 (2010).
56. McGuire, J. T., Nassar, M. R., Gold, J. I. & Kable, J. W. Functionally dissociable influences on learning rate in a dynamic environment. *Neuron* **84**, 870–881 (2014).
57. Costa, V. D., Tran, V. L., Turchi, J. & Averbeck, B. B. Reversal learning and dopamine: a Bayesian perspective. *J. Neurosci.* **35**, 2407–2416 (2015).
58. Mathys, C., Daunizeau, J., Friston, K. J. & Stephan, K. E. A Bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* **5**, 39 (2011).
59. Piray, P. & Daw, N. D. A simple model for learning in volatile environments. *PLoS Comput. Biol.* **16**, e1007963 (2020).
60. Farashahi, S. et al. Metaplasticity as a neural substrate for adaptive learning and choice under uncertainty. *Neuron* **94**, 401–414.e6 (2017).
61. Nassar, M. R. et al. Rational regulation of learning dynamics by pupil-linked arousal systems. *Nat. Neurosci.* **15**, 1040–1046 (2012).
62. Cazé, R. D. & van der Meer, M. A. A. Adaptive properties of differential learning rates for positive and negative outcomes. *Biol. Cybern.* **107**, 711–719 (2013).
63. Louie, K. & Glimcher, P. W. Efficient coding and the neural representation of value. *Ann. N. Y. Acad. Sci.* **1251**, 13–32 (2012).
64. Dabney, W. et al. A distributional code for value in dopamine-based reinforcement learning. *Nature* **577**, 671–675 (2020).

65. Gershman, S. J. Do learning rates adapt to the distribution of rewards? *Psychonomic Bull. Rev.* **22**, 1320–1327 (2015).
66. Daw, N. D., Kakade, S. & Dayan, P. Opponent interactions between serotonin and dopamine. *Neural Netw.* **15**, 603–616 (2002).
67. Frank, M. J., Seeberger, L. C. & O'Reilly, R. C. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**, 1940–1943 (2004).
68. Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. & Palminteri, S. Behavioural and neural characterization of optimistic reinforcement learning. *Nat. Hum. Behav.* **1**, 0067 (2017).
69. Niv, Y., Edlund, J. A., Dayan, P. & O'Doherty, J. P. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J. Neurosci.* **32**, 551–562 (2012).
70. Rosenbaum, G., Grassie, H. & Hartley, C. A. Valence biases in reinforcement learning shift across adolescence and modulate subsequent memory. *eLife* **11**, e64620 (2022).
71. Chambon, V. et al. Information about action outcomes differentially affects learning from self-determined versus imposed choices. *Nat. Hum. Behav.* **4**, 1067–1079 (2020).
72. Palminteri, S., Lefebvre, G., Kilford, E. J. & Blakemore, S.-J. Confirmation bias in human reinforcement learning: evidence from counterfactual feedback processing. *PLoS Comput. Biol.* **13**, e1005684 (2017).
73. Habicht, J., Bowler, A., Moses-Payne, M. E. & Hauser, T. U. Children are full of optimism, but those rose-tinted glasses are fading — reduced learning from negative outcomes drives hyperoptimism in children. *J. Exp. Psychol. Gen.* **151**, 1843–1853 (2022).
74. Villano, W. J. et al. Individual differences in naturalistic learning link negative emotionality to the development of anxiety. *Sci. Adv.* **9**, eadd2976 (2023).
75. Cools, R. et al. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J. Neurosci.* **29**, 1538–1543 (2009).
76. Michely, J., Eldar, E., Erdman, A., Martin, I. M. & Dolan, R. J. Serotonin modulates asymmetric learning from reward and punishment in healthy human volunteers. *Commun. Biol.* **5**, 812 (2022).
77. Cools, R., Robinson, O. J. & Sahakian, B. Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. *Neuropsychopharmacology* **33**, 2291–2299 (2008).
78. Tanaka, S. C. et al. Serotonin affects association of aversive outcomes to past actions. *J. Neurosci.* **29**, 15669–15674 (2009).
79. den Ouden, H. E. M. et al. Dissociable effects of dopamine and serotonin on reversal learning. *Neuron* **80**, 1090–1100 (2013).
80. Moscarello, J. M. & Hartley, C. A. Agency and the calibration of motivated behavior. *Trends Cogn. Sci.* **21**, 725–735 (2017).
81. Ligneul, R. Prediction or causation? Towards a redefinition of task controllability. *Trends Cogn. Sci.* **25**, 431–433 (2021).
82. Raab, H. A., Foord, C., Ligneul, R. & Hartley, C. A. Developmental shifts in computations used to detect environmental controllability. *PLoS Comput. Biol.* **18**, e1010120 (2022).
83. Ligneul, R., Mainen, Z. F., Ly, V. & Cools, R. Stress-sensitive inference of task controllability. *Nat. Hum. Behav.* **6**, 812–822 (2022).
84. Csifcsák, G., Melsæter, E. & Mittner, M. Intermittent absence of control during reinforcement learning interferes with Pavlovian bias in action selection. *J. Cogn. Neurosci.* **32**, 646–663 (2020).
85. Dorfman, H. M., Bhui, R., Hughes, B. L. & Gershman, S. J. Causal inference about good and bad outcomes. *Psychol. Sci.* **30**, 516–525 (2019).
86. Cohen, A. O., Nussenbaum, K., Dorfman, H. M., Gershman, S. J. & Hartley, C. A. The rational use of causal inference to guide reinforcement learning strengthens with age. *NPJ Sci. Learn.* **5**, 16 (2020).
87. Pulcu, E. & Browning, M. Affective bias as a rational response to the statistics of rewards and punishments. *eLife* **6**, e27879 (2017).
88. Dorfman, H. M. et al. Causal inference gates corticostriatal learning. *J. Neurosci.* **41**, 6892–6904 (2021).
89. O'Doherty, J. et al. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* **304**, 452–454 (2004).
90. Amat, J. et al. Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. *Nat. Neurosci.* **8**, 365–371 (2005).
91. Gershman, S. J., Guitart-Masip, M. & Cavanagh, J. F. Neural signatures of arbitration between Pavlovian and instrumental action selection. *PLoS Comput. Biol.* **17**, e1008553 (2021).
92. Palminteri, S. & Lebreton, M. The computational roots of positivity and confirmation biases in reinforcement learning. *Trends Cogn. Sci.* **26**, 607–621 (2022).
93. Langer, E. J. The illusion of control. *J. Pers. Soc. Psychol.* **32**, 311–328 (1975).
94. Lefebvre, G., Summerfield, C. & Bogacz, R. A normative account of confirmation bias during reinforcement learning. *Neural Comput.* **34**, 307–337 (2022).
95. Huys, Q. J. M. & Dayan, P. A Bayesian formulation of behavioral control. *Cognition* **113**, 314–328 (2009).
96. Schubert, J. A., Jagadish, A. K., Binz, M. & Schulz, E. A rational analysis of the optimism bias using meta-reinforcement learning. In *2023 Conference on Cognitive Computational Neuroscience* 557–559 (2023).
97. Greenough, W. T., Black, J. E. & Wallace, C. S. in *Brain Development and Cognition: A Reader* 2nd ed., 186–216 (Wiley, 2008).
98. Knudsen, E. I. Sensitive periods in the development of the brain and behavior. *J. Cogn. Neurosci.* **16**, 1412–1425 (2004).
99. Gabard-Durnam, L. & McLaughlin, K. A. Sensitive periods in human development: charting a course for the future. *Curr. Opin. Behav. Sci.* **36**, 120–128 (2020).
100. Hensch, T. K. Critical period regulation. *Annu. Rev. Neurosci.* **27**, 549–579 (2004).
101. Takesian, A. E. & Hensch, T. K. Balancing plasticity/stability across brain development. *Prog. Brain Res.* **207**, 3–34 (2013).
102. Fawcett, T. W. & Frankenhuis, W. E. Adaptive explanations for sensitive windows in development. *Front. Zool.* **12**, S3 (2015).
103. Golarai, G. & Ghahremani, D. G. The development of race effects in face processing from childhood through adulthood: neural and behavioral evidence. *Dev. Sci.* **24**, e13058 (2021).
104. Kuhl, P. K. et al. Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **363**, 979–1000 (2008).
105. Lin, W. C., Delevich, K. & Wilbrecht, L. A role for adaptive developmental plasticity in learning and decision making. *Curr. Opin. Behav. Sci.* **36**, 48–54 (2020).
106. Anzures, G. et al. Developmental origins of the other-race effect. *Curr. Dir. Psychol. Sci.* **22**, 173–178 (2013).
107. Kuhl, P. K., Tsao, F.-M. & Liu, H.-M. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc. Natl Acad. Sci. USA* **100**, 9096–9101 (2003).
108. Best, C. T., McRoberts, G. W., LaFleur, R. & Silver-Isenstadt, J. Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant. Behav. Dev.* **18**, 339–350 (1995).
109. Kelly, D. J. et al. The other-race effect develops during infancy: evidence of perceptual narrowing. *Psychol. Sci.* **18**, 1084–1089 (2007).
110. McLaughlin, K. A., Sheridan, M. A. & Lambert, H. K. Childhood adversity and neural development: deprivation and threat as distinct dimensions of early experience. *Neurosci. Biobehav. Rev.* **47**, 578–591 (2014).
111. Ellis, B. J., Sheridan, M. A., Belsky, J. & McLaughlin, K. A. Why and how does early adversity influence development? Toward an integrated model of dimensions of environmental experience. *Dev. Psychopathol.* **34**, 447–471 (2022).
112. Mehta, M. A. et al. Hyporesponsive reward anticipation in the basal ganglia following severe institutional deprivation early in life. *J. Cogn. Neurosci.* **22**, 2316–2325 (2010).
113. Hanson, J. L. et al. Behavioral problems after early life stress: contributions of the hippocampus and amygdala. *Biol. Psychiatry* **77**, 314–323 (2015).
114. Dillon, D. G. et al. Childhood adversity is associated with left basal ganglia dysfunction during reward anticipation in adulthood. *Biol. Psychiatry* **66**, 206–213 (2009).
115. Park, A. T. et al. Early childhood stress is associated with blunted development of ventral tegmental area functional connectivity. *Dev. Cogn. Neurosci.* **47**, 100909 (2021).
116. Marusak, H. A., Hatfield, J. R. B., Thomason, M. E. & Rabinak, C. A. Reduced ventral tegmental area–hippocampal connectivity in children and adolescents exposed to early threat. *Biol. Psychiatry Cognit. Neurosci. Neuroimaging* **2**, 130–137 (2017).
117. Fareri, D. S. et al. Altered ventral striatal-medial prefrontal cortex resting-state connectivity mediates adolescent social problems after early institutional care. *Dev. Psychopathol.* **29**, 1865–1876 (2017).
118. Evans, G. W., Li, D. & Whipple, S. S. Cumulative risk and child development. *Psychol. Bull.* **139**, 1342–1396 (2013).
119. Ellis, B. J., Bianchi, J., Griskevicius, V. & Frankenhuis, W. E. Beyond risk and protective factors: an adaptation-based approach to resilience. *Perspect. Psychol. Sci.* **12**, 561–587 (2017).
120. Frankenhuis, W. E., Panchanathan, K. & Nettle, D. Cognition in harsh and unpredictable environments. *Curr. Opin. Psychol.* **7**, 76–80 (2016).
121. Ellwood-Lowe, M. E., Whitfield-Gabrieli, S. & Bunge, S. A. Brain network coupling associated with cognitive performance varies as a function of a child's environment in the ABCD study. *Nat. Commun.* **12**, 7183 (2021).
122. Amso, D. Neighborhood poverty and brain development: adaptation or maturation, fixed or reversible? *JAMA Netw. Open* **3**, e2024139 (2020).
123. Burk, D. C. & Averbeck, B. B. Environmental uncertainty and the advantage of impulsive choice strategies. *PLoS Comput. Biol.* **19**, e1010873 (2023).
124. Frankenhuis, W. E. & Gopnik, A. Early adversity and the development of explore-exploit tradeoffs. *Trends Cogn. Sci.* **27**, 616–630 (2023).
125. Santarelli, S. et al. Evidence supporting the match/mismatch hypothesis of psychiatric disorders. *Eur. Neuropsychopharmacol.* **24**, 907–918 (2014).
126. Schmidt, M. V. Animal models for depression and the mismatch hypothesis of disease. *Psychoneuroendocrinology* **36**, 330–338 (2011).
127. Humphreys, K. L. et al. Exploration-exploitation strategy is dependent on early experience. *Dev. Psychobiol.* **57**, 313–321 (2015).
128. Harms, M. B., Shannon Bowen, K. E., Hanson, J. L. & Pollak, S. D. Instrumental learning and cognitive flexibility processes are impaired in children exposed to early life stress. *Dev. Sci.* **21**, e12596 (2018).
129. Hanson, J. L. et al. Early adversity and learning: implications for typical and atypical behavioral development. *J. Child Psychol. Psychiatry* **58**, 770–778 (2017).
130. Lloyd, A., McKay, R., Sebastian, C. L. & Balsters, J. H. Are adolescents more optimal decision-makers in novel environments? Examining the benefits of heightened exploration in a patch foraging paradigm. *Dev. Sci.* **24**, e13075 (2021).
131. Kamkar, N. H., Lewis, D. J., van den Bos, W. & Morton, J. B. Ventral striatal activity links adversity and reward processing in children. *Dev. Cogn. Neurosci.* **26**, 20–27 (2017).
132. Smith, K. E. & Pollak, S. D. Early life stress and perceived social isolation influence how children use value information to guide behavior. *Child Dev.* **93**, 804–814 (2022).

133. Gerin, M. I. et al. A neurocomputational investigation of reinforcement-based decision making as a candidate latent vulnerability mechanism in maltreated children. *Dev. Psychopathol.* **29**, 1689–1705 (2017).
134. Zador, A. M. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat. Commun.* **10**, 3770 (2019).
135. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
136. Harhen, N. C. & Bornstein, A. M. Interval timing as a computational pathway from early life adversity to affective disorders. *Top. Cogn. Sci.* **16**, 92–112 (2024).
137. Saxe, A. M., McClelland, J. L. & Ganguli, S. A mathematical theory of semantic development in deep neural networks. *Proc. Natl Acad. Sci. USA* **116**, 11537–11546 (2019).
138. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
139. Andrychowicz, M. et al. Learning to learn by gradient descent by gradient descent. *Adv. Neural Inf. Process. Syst.* **29**, 3988–3996 (2016).
140. Bechtle, S. et al. Meta-learning via learned loss. In *Proc. IEEE International Conference on Pattern Recognition* https://doi.org/10.1109/ICPR48806.2021.9412010 (ICPR, 2021).
141. Sutton, R. S. Adapting bias by gradient descent: an incremental version of delta-bar-delta. *AAAI* **92**, 171–176 (1992).
142. Nichol, A., Achiam, J. & Schulman, J. On first-order meta-learning algorithms. Preprint at https://doi.org/10.48550/arXiv.1803.02999 (2018).
143. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
144. Xu, Z. et al. Meta-gradient reinforcement learning with an objective discovered online. *Adv. Neural Inf. Proc. Syst.* **33**, 15254–15264 (2020).
145. Ritter, S., Wang, J. X., Kurth-Nelson, Z. & Botvinick, M. Episodic control as meta-reinforcement learning. Preprint at *bioRxiv* https://doi.org/10.1101/360537 (2018).
146. Hattori, R. et al. Meta-reinforcement learning via orbitofrontal cortex. *Nat. Neurosci.* **26**, 2182–2191 (2023).
147. You, K., Long, M., Wang, J. & Jordan, M. I. How does learning rate decay help modern neural networks? Preprint at https://doi.org/10.48550/arXiv.1908.01878 (2019).
148. Frankenhuis, W. E. & Walasek, N. Modeling the evolution of sensitive periods. *Dev. Cogn. Neurosci.* **41**, 100715 (2020).
149. Xu, Z., van Hasselt, H. & Silver, D. Meta-gradient reinforcement learning. Preprint at https://doi.org/10.48550/arXiv.1805.09801 (2018).
150. Zahavy, T. et al. A self-tuning actor-critic algorithm. *Adv. Neural Inf. Process. Syst.* **33**, 20913–20924 (2020).
151. Zheng, Z., Oh, J. & Satinder, S. On learning intrinsic rewards for policy gradient methods. Preprint at https://doi.org/10.48550/arXiv.1804.06459 (2018).
152. Sanders, B. & Becker-Lausen, E. The measurement of psychological maltreatment: early data on the Child Abuse and Trauma Scale. *Child Abuse Negl.* **19**, 315–323 (1995).
153. Rudolph, K. D. et al. Toward an interpersonal life-stress model of depression: the developmental context of stress generation. *Dev. Psychopathol.* **12**, 215–234 (2000).
154. Young, E. S., Frankenhuis, W. E. & Ellis, B. J. Theory and measurement of environmental unpredictability. *Evol. Hum. Behav.* **41**, 550–556 (2020).
155. Roy, D. et al. in *Symbol Grounding and Beyond* (eds. Vogt, P., Sugita, Y., Tuci, E. & Nehaniv, C.) 192–196 (Springer, 2006).
156. Sullivan, J., Mei, M., Perfors, A., Wojcik, E. & Frank, M. C. SAYCam: a large, longitudinal audiovisual dataset recorded from the infant's perspective. *Open Mind* **5**, 20–29 (2021).
157. Ugarte, E. & Hastings, P. Assessing unpredictability in caregiver-child relationships: insights from theoretical and empirical perspectives. *Dev. Psychopathol.* https://doi.org/10.1017/S0954579423000305 (2022).
158. Tamis-LeMonda, C. S., Kuchirko, Y. & Song, L. Why is infant language learning facilitated by parental responsiveness? *Curr. Dir. Psychol. Sci.* **23**, 121–126 (2014).
159. Ainsworth, M. D. S., Bell, S. M. & Stayton, D. F. in *The Integration of a Child into a Social World* (ed. Richards, M. P. M.) 316, 99–135 (Cambridge Univ. Press, 1974).
160. Csikszentmihalyi, M., Larson, R. & Prescott, S. The ecology of adolescent activity and experience. *J. Youth Adolesc.* **6**, 281–294 (1977).
161. Russell, M. A. & Gajos, J. M. Annual research review: ecological momentary assessment studies in child psychology and psychiatry. *J. Child Psychol. Psychiatry* **61**, 376–394 (2020).
162. Heller, A. S. et al. Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nat. Neurosci.* **23**, 800–804 (2020).
163. Saragosa-Harris, N. M. et al. Real-world exploration increases across adolescence and relates to affect, risk taking, and social connectivity. *Psychol. Sci.* **33**, 1664–1679 (2022).
164. Bath, K., Manzano-Nieves, G. & Goodwill, H. Early life stress accelerates behavioral and neural maturation of the hippocampus in male mice. *Horm. Behav.* **82**, 64–71 (2016).
165. Rice, C. J., Sandman, C. A., Lenjavi, M. R. & Baram, T. Z. A novel mouse model for acute and long-lasting consequences of early life stress. *Endocrinology* **149**, 4892–4900 (2008).
166. Ivy, A. S., Brunson, K. L., Sandman, C. & Baram, T. Z. Dysfunctional nurturing behavior in rat dams with limited access to nesting material: a clinically relevant model for early-life stress. *Neuroscience* **154**, 1132–1142 (2008).
167. Goodkin, F. Rats learn the relationship between responding and environmental events: an expansion of the learned helplessness hypothesis. *Learn. Motiv.* **7**, 382–393 (1976).
168. Overmier, J. B., Patterson, J. & Wielkiewicz, R. M. in *Coping and Health* (eds Levine, S. & Ursin, H.) 1–38 (Springer, 1980).
169. Powell, S. B., Newman, H. A., McDonald, T. A., Bugenhagen, P. & Lewis, M. H. Development of spontaneous stereotyped behavior in deer mice: effects of early and late exposure to a more complex environment. *Dev. Psychobiol.* **37**, 100–108 (2000).
170. Marques, J. M. & Olsson, I. A. S. The effect of preweaning and postweaning housing on the behaviour of the laboratory mouse (*Mus musculus*). *Lab. Anim.* **41**, 92–102 (2007).
171. Ivy, A. S. et al. Hippocampal dysfunction and cognitive impairments provoked by chronic early-life stress involve excessive activation of CRH receptors. *J. Neurosci.* **30**, 13005–13015 (2010).
172. Moriceau, S., Shionoya, K., Jakubs, K. & Sullivan, R. M. Early-life stress disrupts attachment learning: the role of amygdala corticosterone, locus ceruleus corticotropin releasing hormone, and olfactory bulb norepinephrine. *J. Neurosci.* **29**, 15745–15755 (2009).
173. Hartley, C. A., Nussenbaum, K. & Cohen, A. O. Interactive development of adaptive learning and memory. *Annu. Rev. Psychol.* **3**, 59–85 (2021).
174. Zhihong Zeng, A. Survey of affect recognition methods: audio, visual, and spontaneous expressions, 2009. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**, 39–58 (2021).
175. Belo, J. P. R., Azevedo, H., Ramos, J. J. G. & Romero, R. A. F. Deep Q-network for social robotics using emotional social signals. *Front. Robot. AI* **9**, 880547 (2022).
176. Qureshi, A. H., Nakamura, Y., Yoshikawa, Y. & Ishiguro, H. Intrinsically motivated reinforcement learning for human–robot interaction in the real-world. *Neural Netw.* **107**, 23–33 (2018).
177. Kuhn, D. A developmental model of critical thinking. *Educ. Res.* **28**, 16–46 (1999).
178. Kuhn, D. *Education for Thinking* (Harvard Univ. Press, 2005).
179. Joshi, S., Li, Y., Kalwani, R. M. & Gold, J. I. Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron* **89**, 221–234 (2016).
180. Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H. & Balsters, J. H. Pupil diameter covaries with BOLD activity in human locus coeruleus. *Hum. Brain Mapp.* **35**, 4140–4154 (2014).
181. Reimer, J. et al. Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nat. Commun.* **7**, 13289 (2016).
182. Bouret, S. & Sara, S. J. Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends Neurosci.* **28**, 574–582 (2005).
183. Cook, J. L. et al. Catecholaminergic modulation of meta-learning. *eLife* **8**, e51439 (2019).
184. Newcombe, N. S. What is neoconstructivism? neoconstructivism. *Child Dev. Perspect.* **5**, 157–160 (2011).
185. Newcombe, N. S. Cognitive development: changing views of cognitive change. *Wiley Interdiscip. Rev. Cogn. Sci.* **4**, 479–491 (2013).
186. Westermann, G. et al. Neuroconstructivism. *Dev. Sci.* **10**, 75–83 (2007).
187. Karmiloff-Smith, A. *Beyond Modularity: A Developmental Perspective on Cognitive Science* (MIT Press, 1995).
188. Johnson, M. H. Functional brain development in infants: elements of an interactive specialization framework. *Child Dev.* **71**, 75–81 (2000).
189. Westermann, G., Sirois, S., Shultz, T. R. & Mareschal, D. Modeling developmental cognitive neuroscience. *Trends Cogn. Sci.* **10**, 227–232 (2006).
190. Mareschal, D. & Shultz, T. R. Generative connectionist networks and constructivist cognitive development. *Cogn. Dev.* **11**, 571–603 (1996).
191. Astle, D. E., Johnson, M. H. & Akarca, D. Toward computational neuroconstructivism: a framework for developmental systems neuroscience. *Trends Cogn. Sci.* **27**, 726–744 (2023).
192. Elman, J. L. Learning and development in neural networks: the importance of starting small. *Cognition* **48**, 71–99 (1993).
193. Munakata, Y. & McClelland, J. L. Connectionist models of development. *Dev. Sci.* **6**, 413–429 (2003).
194. Fahlman, S. E. The recurrent cascade-correlation architecture. *Adv. Neural Inf. Process. Syst.* **3**, 190–196 (1990).
195. Mata, R., Josef, A. K. & Hertwig, R. Propensity for risk taking across the life span and around the globe. *Psychol. Sci.* **27**, 231–243 (2016).
196. Falk, A. et al. Global evidence on economic preferences. *Q. J. Econ.* **133**, 1645–1692 (2018).
197. Kidd, C., Palmeri, H. & Aslin, R. N. Rational snacking: young children's decision-making on the marshmallow task is moderated by beliefs about environmental reliability. *Cognition* **126**, 109–114 (2013).
198. Yanaoka, K. et al. Cultures crossing: the power of habit in delaying gratification. *Psychol. Sci.* **33**, 1172–1181 (2022).
199. Amir, D. et al. The developmental origins of risk and time preferences across diverse societies. *J. Exp. Psychol. Gen.* **149**, 650–661 (2020).
200. Amir, D. & Jordan, M. R. The behavioral constellation of deprivation may be best understood as risk management. *Behav. Brain Sci.* **40**, e316 (2017).
201. Abebe, T. Reconceptualising children's agency as continuum and interdependence. *Soc. Sci.* **8**, 81 (2019).
202. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
203. Nielsen, M., Haun, D., Kärtner, J. & Legare, C. H. The persistent sampling bias in developmental psychology: a call to action. *J. Exp. Child Psychol.* **162**, 31–38 (2017).
204. Tenenbaum, J. B., Kemp, C., Griffiths, T. L. & Goodman, N. D. How to grow a mind: statistics, structure, and abstraction. *Science* **331**, 1279–1285 (2011).
205. Wellman, H. M. & Gelman, S. A. Cognitive development: foundational theories of core domains. *Annu. Rev. Psychol.* **43**, 337–375 (1992).

206. Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. Building machines that learn and think like people. *Behav. Brain Sci.* **40**, e253 (2017).
207. Nettle, D., Frankenhuis, W. E. & Rickard, I. J. The evolution of predictive adaptive responses in human life history. *Proc. Biol. Sci.* **280**, 20131343 (2013).
208. Gogtay, N. et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl Acad. Sci. USA* **101**, 8174–8179 (2004).
209. Averbeck, B. B. Pruning recurrent neural networks replicates adolescent changes in working memory and reinforcement learning. *Proc. Natl Acad. Sci. USA* **119**, e2121331119 (2022).
210. Ajemian, R., D'Ausilio, A., Moorman, H. & Bizzi, E. A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits. *Proc. Natl Acad. Sci. USA* **110**, E5078–E5087 (2013).
211. Yamins, D. L. K. & DiCarlo, J. J. Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* **19**, 356–365 (2016).
212. Findling, C. & Wyart, V. Computation noise promotes cognitive resilience to adverse conditions during decision-making. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.10.145300 (2020).
213. Plappert, M. et al. Parameter space noise for exploration. Preprint at:*arXiv* https://doi.org/10.48550/arXiv.1706.01905 (2017).
214. Fortunato, M. et al. Noisy networks for exploration. In *Proc. International Conference on Learning Representations (ICLR)* (2018).
215. McIntosh, A. R. et al. The development of a noisy brain. *Arch. Ital. Biol.* **148**, 323–337 (2010).
216. Smith, L. B., Jayaraman, S., Clerkin, E. & Yu, C. The developing infant creates a curriculum for statistical learning. *Trends Cogn. Sci.* **22**, 325–336 (2018).
217. Kidd, C. & Hayden, B. Y. The psychology and neuroscience of curiosity. *Neuron* **88**, 449–460 (2015).
218. Gottlieb, J., Oudeyer, P.-Y., Lopes, M. & Baranes, A. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends Cogn. Sci.* **17**, 585–593 (2013).
219. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. Curriculum learning. In *Proc. 26th Annual International Conference on Machine Learning* 41–48 (Association for Computing Machinery, 2009).
220. Oudeyer, P.-Y. & Kaplan, F. What is intrinsic motivation? A typology of computational approaches. *Front. Neurorobot.* **1**, 6 (2007).
221. Forestier, S., Mollard, Y. & Oudeyer, P.-Y. Intrinsically motivated goal exploration processes with automatic curriculum learning. *J. Mach. Learn. Res.* **23**, 1–41 (2022).

Acknowledgements

The authors thank Bruno Averbeck, Rheza Budiono, Nathaniel Daw, Nora Harhen and Akshay Jagadish for helpful feedback on the manuscript. Preparation of this manuscript was supported by the C.V. Starr Fellowship (to K.N.).

Author contributions

All authors conceptualized the article content. K.N. wrote the article draft and all authors edited the manuscript before submission.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Psychology* thanks Dorsa Amir, who co-reviewed with Annya Dahmani; Jane Wang; and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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