Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

What do reinforcement learning models measure? Interpreting model parameters in cognition and neuroscience

Maria K Eckstein[1], Linda Wilbrecht[1,2] and Anne GE Collins[1,2]

Reinforcement learning (RL) is a concept that has been invaluable to fields including machine learning, neuroscience, and cognitive science. However, what RL entails differs between fields, leading to difficulties when interpreting and translating findings. After laying out these differences, this paper focuses on cognitive (neuro)science to discuss how we as a field might overinterpret RL modeling results. We too often assume — implicitly — that modeling results *generalize* between tasks, models, and participant populations, despite negative empirical evidence for this assumption. We also often assume that parameters measure specific, unique (neuro) cognitive processes, a concept we call *interpretability*, when evidence suggests that they capture different functions across studies and tasks. We conclude that future computational research needs to pay increased attention to implicit assumptions when using RL models, and suggest that a more systematic understanding of contextual factors will help address issues and improve the ability of RL to explain brain and behavior.

Addresses
[1] Department of Psychology, UC Berkeley, 2121 Berkeley Way West, Berkeley 94720, CA, USA

[2] Helen Wills Neuroscience Institute, UC Berkeley, 175 Li Ka Shing Center, Berkeley 94720, CA, USA

Corresponding author: Collins, Anne GE (annecollins@berkeley.edu)

Current Opinion in Behavioral Sciences 2021, **41C**:128-137

This review comes from a themed issue on **Cognition and Perception — 'Value-based decision-making'**

Edited by **Bernard Balleine** and **Laura Bradfield**

https://doi.org/10.1016/j.cobeha.2021.06.004

Introduction

Reinforcement learning (RL) is an exploding field. In the domain of machine learning, it has led to tremendous progress in the last decade, ranging from the creation of artificial agents that can beat humans at complex games, such as Go [1] and StarCraft [2], to successful deployment of internet balloons in the stratosphere [3]. In cognitive neuroscience, RL models have been used successfully to capture a broad range of latent learning-related phenomena, at the level of both behavior [4,5] and neural signals [6]. However, the impression that RL can help us identify reasonable and predictive latent variables hides heterogeneity in what RL variables reflect, even within cognitive neuroscience. The success of RL has fed a notion of omniscience that RL can peer into the brain and behavior and surgically isolate and measure essential functions. As this notion grows with the popular uptake of RL methods, it sometimes leads to overgeneralization and overinterpretation of findings.

Here, we argue that a more nuanced view is better supported empirically and theoretically. We first discuss how RL is used in distinct subfields, highlighting shared and distinct components. Then, we examine where cognitive neuroscience may be overstepping in its interpretation, and conclude that, when properly contexualized, RL models retain great value for the field.

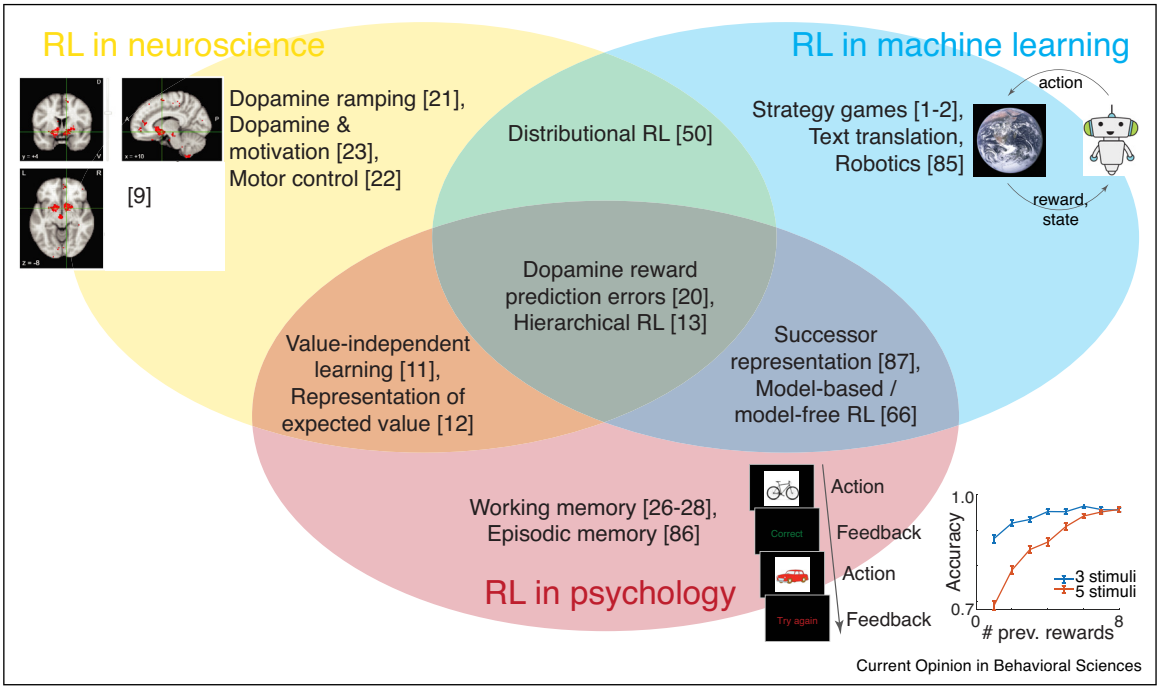
RL in machine learning, psychology, and neuroscience

In machine learning, RL is defined as a class of learning problems and a family of algorithms that solve these problems. An RL agent can be in any of a set of states, take actions to change states, and receive rewards/punishments (Figure 1, top-right). RL agents are designed to optimize a specific objective: the expected sum of discounted future rewards. A wide family of RL algorithms offers solutions that achieve this objective [10], for example model-free RL, which estimates the values of actions based on reward prediction errors (Figure 2a, top).

In psychology, RL defines a psychological process and a method for its study. RL occurs when an organism learns to make choices (or predict outcomes) directly based on experienced rewards/punishments (rather than indirectly through instructions, for example). This includes simple situations, such as those historically studied by behaviorists (classical [6,11•] and instrumental conditioning [12]), as well as more complex ones, such as learning over longer time horizons [13,14], meta-learning [15], and learning across multiple contexts [16,17].

Neuroscientists investigating RL usually focus on a well-defined network of regions that implements value

Figure 1



The meaning of 'RL' differs between neuroscience, machine learning, and psychology, reflecting a specific brain network, a family of problems and algorithms, and a type of learning, respectively. The concepts are related: RL models successfully capture aspects of RL behavior and brain signals, and some RL behaviors rely on the RL brain network. The dopamine reward prediction error hypothesis combines ideas from all three fields. However, there are also significant discrepancies in what RL means across fields, such that activity in the brain's RL network might not relate to RL behavior and might not be captured by RL models (e.g. dopamine ramping in neuroscience). Importantly, RL behavior may rely on non-RL brain systems and may or may not be captured by RL algorithms. Recent trends have aimed to increase communication between fields and emphasize areas of mutual benefits [7•,8]. RL in neuroscience inset shows the neurosynth automated meta-analysis for 'reinforcement learning' ($x = 10, y = 4, z = -8$), highlighting striatal function [9]. RL in cognition inset shows that participants become more likely to select a rewarded choice the more previous rewards they have experienced (data replotted from [5]). RL in machine learning shows the agent-environment loop at the basis of RL theory [10,85–87].

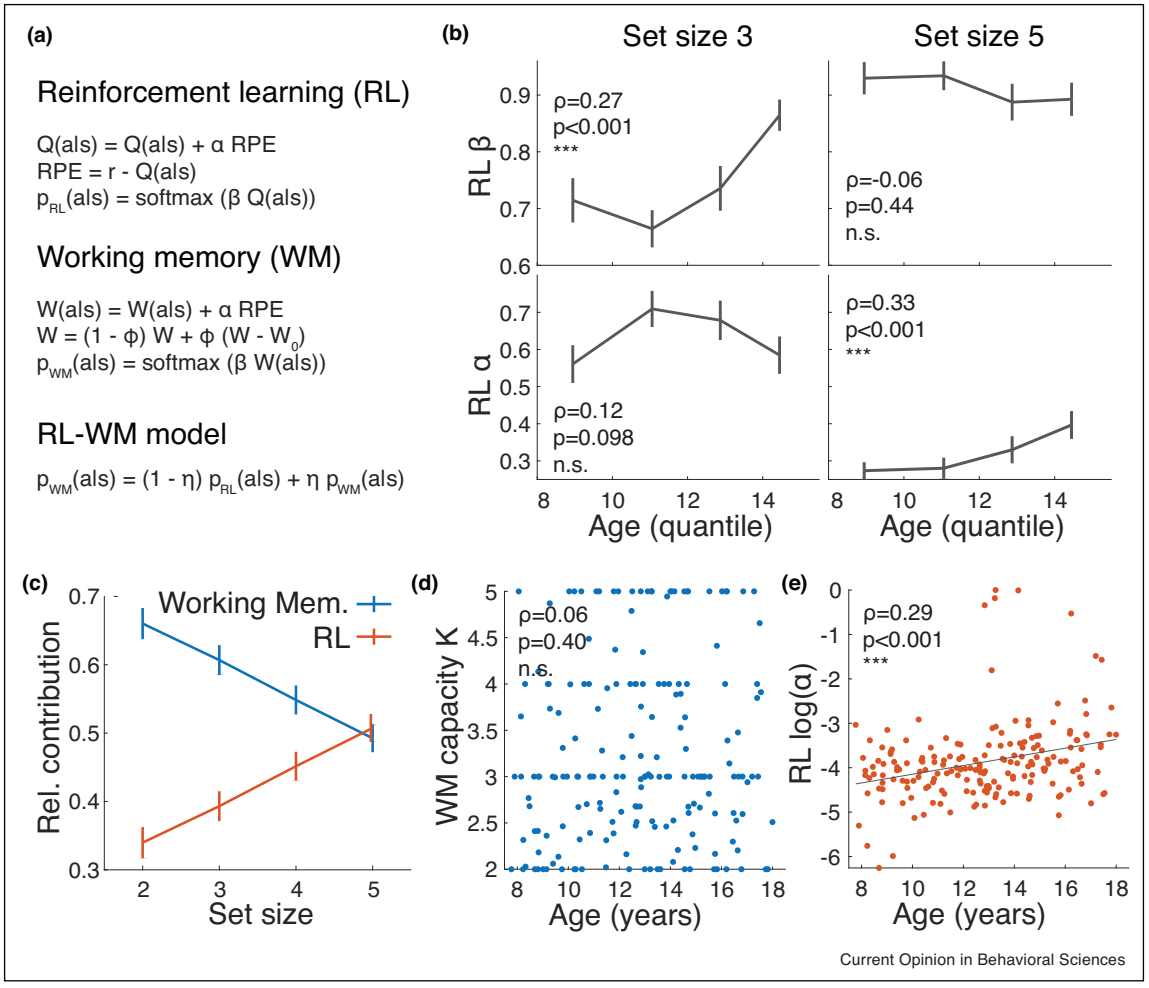
learning. These include cortico-basal-ganglia loops, and in particular the striatum (Figure 1a), thought to encode RL values, and dopamine neurons, thought to signal temporal-difference reward-prediction errors (*RPEs*; Figure 2a) [6,9,18–21].

The meaning of 'RL' overlaps in these three communities (Figure 1), and RL algorithms from AI have been successful at capturing biological RL behavior and neural function. However, there are also important discrepancies. For example, many functions of the brain's RL network do not relate to RL behavior, such as dopamine's role in motor control [22] or cognitive effort [23]. On the other hand, some RL brain functions that do relate to RL behavior are poorly explained by classic RL models, such as dopamine's role in value-independent learning [11•]. Furthermore, many aspects of learning from reward do not depend on the brain's RL network, whether they are captured by RL algorithms or not. For example, hippocampal episodic memory [24,25] and prefrontal working memory [26,27•,28] contribute to RL behavior, but are often not explicitly modeled in RL, obscuring the contribution of non-RL neural processes to learning.

Because of these differences in meaning, the term 'RL' can cause ambiguity and lead to misinterpretations. Figure 2 provides an example in which an RL model leads to conflicting conclusions as to how RL parameters change with age when applied to two slight variants of the same task. This conflict is reconciled, however, by recognizing that working memory contributes most learning in one variant, whereas RL does in the other [5].

Because RL's meaning is ambiguous, it is often unclear how RL model variables (e.g. parameters such as learning rates or decision noise; reward prediction errors; RL values) should be interpreted in models of human and animal learning. In the following, we show that the field often optimistically assumes that model variables are readily interpretable and naturally generalize between studies. We then show that these beliefs are oftentimes not well supported, and offer an alternative interpretation.

Figure 2



Fitting standard RL models can lead to the wrong impression that cognitive processing is purely based on RL. **(a)** Update equations for the RL-WM model. $Q(a|s)$ indicates the RL state-action value of action a in state s, which is updated based on the reward prediction error RPE. $W(a|s)$ is the working-memory weight of a in s, and ϕ is a forgetting parameter, β is the decision noise, and η the mixing parameter combining RL and working memory processes. For model details, see [5,29]. **(b)** When separate standard RL models are fit to different contexts within the same task (here, the number of stimuli [6]), they provide different answers as to how age affects RL model parameters (decision noise β, top; learning rate α, bottom). Contexts with fewer stimuli ('Set size 3', left) suggest that age does not affect learning rates, whereas contexts with more stimuli ('Set size 5', right) suggest that learning rates increase with age. Inset statistics show non-parametric Spearman correlation coefficients ρ and P-values ($N = 187$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (c)–(e) When using a model that fits all contexts jointly by combining RL processes with working memory ('RL-WM' model), these discrepancies are resolved [5]. **(c)** The RL-WM model reveals that the relative contributions of RL compared to working memory η differ between contexts. A standard RL model would falsely attribute working-memory processes in contexts with small sizes to the RL system, in this example suggesting that learning rates do not change with age (a, set size 3). **(d)** Working memory capacity in the RL-WM model was not related to participants' ages, explaining why learning rates did not increase with age in (A, set size 3), in which working memory contributed most to learning. **(e)** RL learning rates in the RL-WM model increased with age. Since RL contributed more to learning in set size 5 (c), this was detected in the standard RL model of only set size 5 (b). Data reanalyzed from [5].

Interpretability and generalizability of RL model variables

What do 'cognitive' models measure?

RL models attempt to approximate behavior by fitting free parameters [30–33], and are used by most researchers to elucidate cognitive and/or neural function (Box 1): RL has emerged as a key framework for modeling and understanding *decision-making*[3] [34]. The reason why models of behavior are used as 'cognitive models' is that they implement hypotheses about cognition. Therefore, the good fit of a model to behavior implies that participants could have employed the modeled algorithm

[3] Emphasis added.

Box 1 Representative statements from the literature that imply interpretability and generalizability[4]

- **Interpretability**: Computational models have been described as 'illuminating [. . .] *cognitive processes* or *neural representations* that are otherwise difficult to tease apart' [37••]; clarifying 'the *neural processes* underlying *decision-making*' [18]; and revealing 'what computations are performed in *neuronal populations* that support a particular *cognitive process*'[5] [38]. This highlights the common assumption that computational models can reveal cognitive and neural processes and identify specific, 'theoretically meaningful' [39] elements of (neuro)cognitive function.Models are thereby often expected to provide the 'linking propositions' [40] between cognition and neural function, '*mapping* latent decision-making processes onto dissociable neural substrates' [41••] and '*link[ing]* cognitive mechanisms to [clinical] symptoms' [38].These links are often assumed to be specific and one-to-one: 'Dopamine neurons code an error in the prediction of reward' [20]; 'corticostriatal loops enable state-dependent value-based choice' [27•]; 'striatal areas [. . .] support reinforcement learning, and frontoparietal attention areas [. . .] support executive control processes' [42]; 'individual differences in DA clearance and frontostriatal coordination may serve as markers for RL' [43]; and 'BOLD activity in the VS, dACC, and vmPFC is correlated with learning rate, expected value, and prediction error, respectively' [44]. This shows that computational variables are often interpreted as specific (neuro)cognitive functions, revealing an assumption of *interpretability*.
- **Generalizability**: Empirical parameter distributions obtained in one task were described as 'fairly *transferable*' [45] and used as priors when fitting parameters to a new task [46], revealing the belief that model parameters generalize between studies, tasks, and models.

Many have aimed to find regularities in parameter findings between studies: '[D]ifferential learning rates *tend to be* biased in the direction of learning from positive RPEs' [47]; 'this finding *[supports] previous results* on decreased involvement of the reinforcement learning system when cortical resources [. . .] support task execution' [42]; from our own work: 'there was [. . .] a bias towards learning from positive feedback, which is *consistent with other work*' [5].

cognitively. Nevertheless, stronger conclusions are often drawn: For example, the good fit of inference algorithms to human behavior and brain function has been taken as evidence that human brains implement inference [17]. However, there always is an infinite number of alternative algorithms that would fit behavior equally well, such that inferring participants' cognitive algorithms through model fitting is impossible [33,35,36•].

Interpretability and generalizability

This notion that computational models — astonishingly — isolate and measure intrinsic (neuro)cognitive processes from observable behavior has contributed to their attractiveness as a research method. However, we believe we need to temper our optimism in two areas: *interpretability* and *generalizability* (Figure 3).

Interpretability means that model variables (e.g. parameters, reward prediction errors) isolate specific, fundamental, and invariant elements of (neuro)cognitive processing: Decomposing behavior into model variables is seen as a way of carving cognition at its joints, producing model variables that are of essential nature. Generalizability means that model variables capture inherent individual characteristics (e.g. a person with a high learning rate), such that we can robustly infer the same parameter for the same person across different contexts, tasks, and model variants.

Though rarely stated explicitly, assumptions about interpretability and generalizability lie at the heart of much current computational cognitive research (including our own), as we show in the literature survey above (Box 1), and play a consequential role in interpreting and guiding future research. However, we also show that empirical support for interpretability and generalizability is ambivalent at best, and often negative. We highlight a recent multi-task within-participants study from our group that explores precisely when model parameters do and do not generalize between tasks, and how dissimilar the cognitive processes are they capture (interpretability).

Interpretability

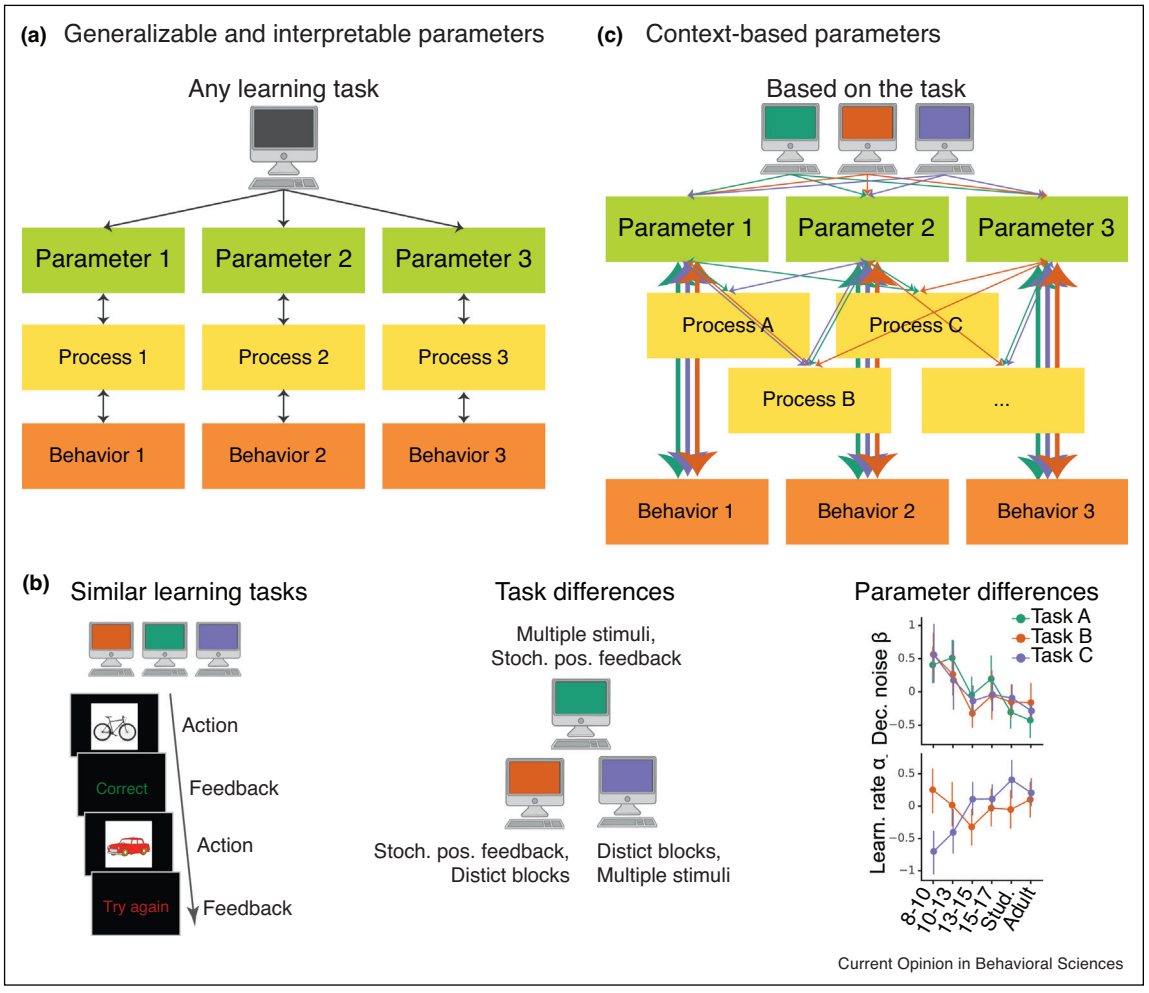
Many research practices are deeply invested in the interpretability of RL (Box 1). The computational neurosciences, for example, aim to link computational variables to specific neural functions, searching for one-to-one mappings that would allow the inference of one from the other [6,12,43,48]. Prominent examples of *interpretable* mappings are the links between the midbrain-dopamine system and RL reward prediction errors [20,49,50], and between striatal function and value learning [19,51–53]. Computational psychiatry aims to map model variables onto psychiatric diagnoses or symptoms, in an effort to obtain diagnostic tools and causal explanations of aberrant processing [38,39,41••,54]. Developmental research aims to map age-related changes in model variables onto developing neural function and real-world behavior [37••,55,56]. In sum, the conviction in model interpretability is evident in the practice of interpreting model variables as specific cognitive processes, unique neural substrates, and well-delineated psychiatric symptoms.

Generalizability

Assumptions about parameter generalizability are also widespread. In computational neuroscience, model variables are routinely expected to measure the same latent neural substrates, even when the underlying task, model, or participant samples differ [18–20,57–60]. For example, fields studying individual differences, such as clinical [38,39] and developmental psychology [37••,55,56], aim

[4] We acknowledge that these statements may not represent the full complexity of researchers' knowledge, as many are aware of modeling limitations.

Figure 3



What model variables (e.g. parameters) measure in psychology and neuroscience. **(a)** View based on interpretability and generalizability. In this view — implicitly taken by much current research — models are fitted in order to reveal individuals' intrinsic characteristics, whereby model parameters reflect clearly delineated, separable, and unique (neuro)cognitive processes. This concept of *interpretability* is shown in the figure in that every model parameter captures one specific cognitive process (bidirectional arrows between parameter and process), and that cognitive processes are separable from each other (no connections between processes). Specific task characteristics are neglected as irrelevant, a concept we call *generalizability*, which is evident in that parameters of 'any learning task' (within reason) are expected to capture the same cognitive processes. **(b)** In our empirical study [76••], participants worked on three learning tasks with similar structure (left), but slight differences (middle), reflecting differences in the literature. We created three RL models that captured the behavior in each task [4,5,69]. Compared between tasks but within participants, learning rate parameters showed poor interpretability and generalizability [76••]: Both absolute values and age trajectories (right, bottom) differed vastly, and individual differences in one task could not be predicted by those in other tasks, as would be expected if they were interpretable as the same (neuro)cognitive substrate. Other parameters, most notably decision noise (right, top), were more generalizable and interpretable, in accordance with emerging patterns in the literature [37••], even though they also lacked a shared core of variance across tasks (more for more dissimilar tasks). In contrast, the mappings between parameters and behavioral features were consistent across tasks, suggesting that parameters generalized in terms of behavioral processes, but not cognitive ones. **(c)** Updated view that acknowledges the role of context in computational modeling (e.g. task characteristics, model parameterization, participant characteristics). Which cognitive processes are captured by each model parameter is influenced by the task (green, orange, blue), as shown by distinct connections between parameters and cognitive processes. Different parameters within the same task can capture overlapping cognitive processes (not interpretable), and the same parameters can capture different processes depending on the task (not generalizable). However, parameters likely capture consistent behavioral patterns across tasks (thick vertical arrows).

to identify how model variables covary with other variables of interest (e.g. age, traits, symptoms) in a systematic way across studies, and review articles and discussion sections commonly compare modeling variables between studies.

Evidence against interpretability and generalizability

However, meta-reviews suggest that interpretability and generalizability might be overassumptions, common in classic psychological research [61••] and RL modeling

[41••]. RL appears interpretable because multiple studies have replicated mappings between RL variables and specific neural function. However, these mappings are not as consistent as expected: The famous mapping between dopamine/striatal activity and reward prediction errors, for example, supported by classic and recent research [6,20], varies considerably between studies based on details of the experimental protocol, as shown in several recent meta-analyses [57,59,62]. Discrepancies are also evident in the mapping between RL variables and cognitive function. For example, learning rates are often interpreted as incremental updating (dopamine-driven neural plasticity) in classical conditioning [20], but also as reward sensitivity [63], sampling from (hippocampal) episodic memory [25], the ability to optimally weigh decision outcomes [64], or approximate inference [4], in other tasks. There is substantial variance between studies in terms of which neural and which cognitive processes underlie the same RL variables, contradicting the notion of interpretability.

Evidence for generalizability is also weak: Similar adult samples have differed strikingly in terms of their average estimated RL learning rates (0.05–0.7) [44,63,65,66] and 'positivity bias' [47,67,68•], depending on the underlying task and model parameterization. In developmental samples, the trajectories of RL learning rates have shown increases [5,63,69], decreases [70], U-shaped trajectories [4], or no change [71] in the same age range. Similar discrepancies have also arisen in the computational psychiatry literature [38,39,72,73]. These inconsistencies would not be expected if model variables were an inherent property of participants that could be assessed independently of study specifics, that is, if models were generalizable.

Many in our community have noticed such discrepancies and invoked methodological differences between studies to explain them [12,37••,44,62,74,75]. However, this insight has rarely been put into practice, and model variables keep being compared between studies (Box 1). To remedy this, we assessed interpretability and generalizability empirically, comparing RL parameters from three tasks performed by the same subjects in a developmental sample (291 subjects aged 8–30; Figure 3b) [4,5,69,76••]. We found generalizability but poor interpretability for decision noise, and a fundamental lack of both interpretability and generalizability for learning rates (Figure 3c).

A likely reason why generalizability and interpretability are lacking in many cases is that computational models are fundamentally models of behavior, and not cognition. Because participants — reasonably — behave differently in different tasks (e.g. repeating non-rewarded actions in stochastic tasks, but not in deterministic ones [76••]), estimated parameters (e.g. learning rates) differ as well. Such differences do not necessarily reflect a failure of computational models to measure intrinsic processes, but likely the fact that the same parameters capture different behaviors and different cognitive processes when applied to different tasks (Figure 3b,c) [76••].

Another reason for lacking generalizability and interpretability is that the design of computational models, a researcher degree of freedom [35,36•], can impact parameters severely, as recent research has highlighted [47, 67,68•]. Because the same models can be parameterized differently [77], and models with different equations can approximate similar processes [4], model differences are a ubiquitous feature of computational modeling.

To explain parameter discrepancies, others have argued that participants adapt their parameter values to tasks based on optimality [37••], or that task characteristics (e.g. uncertainty) influence neural processing (e.g. dopamine function), which is reflected in differences in model variables (e.g. reward prediction errors) [78,79]. Whether choices are aligned with participants' goals also fundamentally impacts neural RL processes [80•], and so do other common task characteristics [59]. This shows that small task differences impact behavior, neural processing, and computational variables. Even though RL models might successfully capture behavior in each task, parameters likely capture different aspects each time, leading to a lack of interpretability and generalizability.

Conclusion and outlook

A tremendous literature has shown RL's potential and successes — this opinion piece emphasizes some caveats, showing that RL is not a single concept and that RL models are a broad family that reflects a range of cognitive and neural processes.

A lack of interpretability and generalizability has major implications for the comparison of model variables between tasks, a practice that forms the basis for many review articles, meta-analyses, introduction and discussion sections of empirical papers, and for directing future research. Evidence suggests that in many cases, parameters cannot directly be compared between studies, and capture different (neuro)cognitive processes depending on task characteristics. Future research needs to determine which model variables do and do not generalize, over which domain, and what the determining factors are. In the meantime, researchers should be more nuanced when comparing results between studies, and acknowledge contextual factors that might limit generalizability. Lastly, what model variables measure might differ for each task, and researchers should validate variables on a task-by-task basis, relating them to behavioral measures or individuals' traits, and using simulations to determine their precise role in specific tasks.

Another solution is to explicitly model variability between features that should be generalized over, including task characteristics (Figure 2), models, participants, and potentially even neural processes [61••]. Several studies have made strides in this direction, incorporating features that are intrinsic to participants (e.g., working memory [5,29], attention [28], development [37••,81,56]), or extrinsic (e.g., task time horizon [13,14], context changes [16]), thus broadening the domain over which models generalize. However, infinitely many features likely affect RL processes, rendering entirely general models infeasible. Researchers therefore need to select a domain of interest for each model, and acknowledge this choice. As authors, reviewers, and editors, we should balance our excitement about general statements with our knowledge about the inherent limitations of all models, including RL. Future research needs to determine whether similar issues arise for other model families, such as sequential sampling [82,83], Bayesian inference [4,28,84], and others.

We hope that this explicit discussion of assumptions and overassumptions will help our field solve the mysteries of the brain as modeling — with its limitations — is embraced by a growing audience.

Conflict of interest statement

Nothing declared.

Acknowledgments

This work was in part supported by National Science Foundation grant 1640885 SL-CN: Science of Learning in Adolescence to AGEC and LW, NIH grant 1U19NS113201 to LW, and NIMHRO1MH119383 to AGEC.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, van den Driessche G, Graepel T, Hassabis D: **Mastering the game of Go without human knowledge**. *Nature* 2017, **550**:354-359 http://dx.doi.org/10.1038/nature24270.

2. Vinyals O, Babuschkin I, Czarnecki WM, Mathieu M, Dudzik A, Chung J, Choi DH, Powell R, Ewalds T, Georgiev P, Oh J, Horgan D, Kroiss M, Danihelka I, Huang A, Sifre L, Cai T, Agapiou JP, Jaderberg M, Vezhnevets AS, Leblond R, Pohlen T, Dalibard V, Budden D, Sulsky Y, Molloy J, Paine TL, Gulcehre C, Wang Z, Pfaff T, Wu Y, Ring R, Yogatama D, Wünsch D, McKinney K, Smith O, Schaul T, Lillicrap T, Kavukcuoglu K, Hassabis D, Apps C, Silver D: **Grandmaster level in StarCraft II using multi-agent reinforcement learning**. *Nature* 2019, **575**:350-354 http://dx.doi.org/10.1038/s41586-019-1724-z.

3. Bellemare MG, Candido S, Castro PS, Gong J, Machado MC, Moitra S, Ponda SS, Wang Z: **Autonomous navigation of stratospheric balloons using reinforcement learning**. *Nature* 2020, **588**:77-82 http://dx.doi.org/10.1038/s41586-020-2939-8.

4. Eckstein MK, Master SL, Dahl RE, Wilbrecht L, Collins AGE: **Understanding the unique advantage of adolescents in stochastic, volatile environments: combining reinforcement learning and Bayesian Inference**. *bioRxiv* 2020 http://dx.doi.org/10.1101/2020.07.04.187971.

5. Master SL, Eckstein MK, Gotlieb N, Dahl R, Wilbrecht L, Collins AGE: **Disentangling the systems contributing to changes in learning during adolescence**. *Dev Cogn Neurosci* 2020, **41**:100732 http://dx.doi.org/10.1016/j.dcn.2019.100732.

6. Maes EJP, Sharpe MJ, Usypchuk AA, Lozzi M, Chang CY, Gardner MPH, Schoenbaum G, Iordanova MD: **Causal evidence supporting the proposal that dopamine transients function as temporal difference prediction errors**. *Nat Neurosci* 2020, **23**:176-178 http://dx.doi.org/10.1038/s41593-019-0574-1.

7. • Neftci EO, Averbeck BB: **Reinforcement learning in artificial and biological systems**. *Nat Mach Intell* 2019, **1**:133-143 http://dx.doi.org/10.1038/s42256-019-0025-4
A review that highlights the strengths of, differences between, and areas of potential mutual benefit for artificial intelligence and cognitive neuroscience.

8. Collins AGE: **Reinforcement learning: bringing together computation and cognition**. *Curr Opin Behav Sci* 2019, **29**:63-68 http://dx.doi.org/10.1016/j.cobeha.2019.04.011.

9. Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD: **Large-scale automated synthesis of human functional neuroimaging data**. *Nat Methods* 2011, **8**:665-670 http://dx.doi.org/10.1038/nmeth.1635.

10. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. edn 2nd. Cambridge, MA; London, England: MIT Press; 2017.

11. • Sharpe MJ, Batchelor HM, Mueller LE, Yun Chang C, Maes EJP, Niv Y, Schoenbaum G: **Dopamine transients do not act as model-free prediction errors during associative learning**. *Nat Commun* 2020, **11**:106 http://dx.doi.org/10.1038/s41467-019-13953-1
In three elegant experiments, the authors show that optogenetically activating dopamine neurons in the VTA induces learning, but not by affecting values, as expected based on RL theory.

12. Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD: **Dissociable dopamine dynamics for learning and motivation**. *Nature* 2019, **570**:65-70 http://dx.doi.org/10.1038/s41586-019-1235-y.

13. Botvinick M: **Hierarchical reinforcement learning and decision making**. *Curr Opin Neurobiol* 2012, **22**:956-962 http://dx.doi.org/10.1016/j.conb.2012.05.008.

14. Xia L, Collins AGE: **Temporal and state abstractions for efficient learning, transfer and composition in humans**. *Psychol Rev* 2021.

15. Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M: **Prefrontal cortex as a meta-reinforcement learning system**. *Nat Neurosci* 2018, **21**:860-868 http://dx.doi.org/10.1038/s41593-018-0147-8.

16. Eckstein MK, Collins AGE: **Computational evidence for hierarchically structured reinforcement learning in humans**. *Proc Natl Acad Sci U S A* 2020, **117**:29381-29389 http://dx.doi.org/10.1073/pnas.1912330117.

17. Findling C, Chopin N, Koechlin E: **Imprecise neural computations as a source of adaptive behaviour in volatile environments**. *Nat Hum Behav* 2021, **5**:99-112 http://dx.doi.org/10.1038/s41562-020-00971-z.

18. Niv Y: **Reinforcement learning in the brain**. *J Math Psychol* 2009, **53**:139-154.

19. Frank MJ, Claus ED: **Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal**. *Psychol Rev* 2006, **113**:300-326 http://dx.doi.org/10.1037/0033-295X.113.2.300.

20. Schultz W, Dickinson A: **Neuronal coding of prediction errors**. *Annu Rev Neurosci* 2000, **23**:473-500 http://dx.doi.org/10.1146/annurev.neuro.23.1.473.

21. Wang Y, Toyoshima O, Kunimatsu J, Yamada H, Matsumoto M: **Tonic firing mode of midbrain dopamine neurons continuously**

tracks reward values changing moment-by-moment. *eLife* 2021 http://dx.doi.org/10.7554/eLife.63166.

22. Meder D, Herz DM, Rowe JB, Lehéricy S, Siebner HR: **The role of dopamine in the brain — lessons learned from Parkinson's disease**. *NeuroImage* 2019, **190**:79-93 http://dx.doi.org/10.1016/j.neuroimage.2018.11.021.

23. Westbrook A, Bosch Rvd, Määttä JI, Hofmans L, Papadopetraki D, Cools R, Frank MJ: **Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work**. *Science* 2020, **367**:1362-1366 http://dx.doi.org/10.1126/science.aaz5891.

24. Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D, Burgess N, Daw ND: **Hippocampal contributions to model-based planning and spatial memory**. *Neuron* 2019, **102**:683-693.e4 http://dx.doi.org/10.1016/j.neuron.2019.02.014.

25. Bornstein AM, Norman KA: **Reinstated episodic context guides sampling-based decisions for reward**. *Nat Neurosci* 2017, **20**:997-1003 http://dx.doi.org/10.1038/nn.4573.

26. Collins AGE, Frank MJ: **Within- and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory**. *Proc Natl Acad Sci U S A* 2018, **115**:2502-2507 http://dx.doi.org/10.1073/pnas.1720963115.

27. • Rmus M, McDougle S, Collins AGE: **The role of executive function in shaping reinforcement learning**. *Curr Opin Behav Sci* 2021, **38**:66-73 http://dx.doi.org/10.1016/j.cobeha.2020.10.003

A recent review article that highlights the contributions of executive function to RL processes.

28. Radulescu A, Niv Y, Ballard I: **Holistic reinforcement learning: the role of structure and attention**. *Trends Cogn Sci* 2019, **23**:278-292 http://dx.doi.org/10.1016/j.tics.2019.01.010.

29. Collins AGE, Frank MJ: **How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis: Working memory in reinforcement learning**. *Eur J Neurosci* 2012, **35**:1024-1035 http://dx.doi.org/10.1111/j.1460-9568.2011.07980.x.

30. Wilson RC, Collins AG: **Ten simple rules for the computational modeling of behavioral data**. *eLife* 2019, **8**:e49547 http://dx.doi.org/10.7554/eLife.49547.

31. Palminteri S, Wyart V, Koechlin E: **The importance of falsification in computational cognitive modeling**. *Trends Cogn Sci* 2017, **21**:425-433 http://dx.doi.org/10.1016/j.tics.2017.03.011.

32. Guest O, Martin AE: **How computational modeling can force theory building in psychological science**. *Perspect Psychol Sci* 2021 http://dx.doi.org/10.1177/1745691620970585.

33. Blohm G, Kording KP, Schrater PR: **A how-to-model guide for neuroscience**. *eNeuro* 2020, **7** http://dx.doi.org/10.1523/ENEURO.0352-19.2019.

34. Diuk C, Schapiro A, Córdova N, Ribas-Fernandes J, Niv Y, Botvinick M: **Divide and conquer: hierarchical reinforcement learning and task decomposition in humans**. *Computational and Robotic Models of the Hierarchical Organization of Behavior*. Berlin, Heidelberg: Springer; 2013, 271-291 http://dx.doi.org/10.1007/978-3-642-39875-9_12.

35. Uttal WR: **On some two-way barriers between models and mechanisms**. *Percept Psychophys* 1990, **48**:188-203 http://dx.doi.org/10.3758/BF03207086.

36. • Navarro DJ: **Between the devil and the deep blue sea: tensions between scientific judgement and statistical model selection**. *Comput Brain Behav* 2019, **2**:28-34 http://dx.doi.org/10.1007/s42113-018-0019-z

An insightful commentary that highlights many of the issues that arise during statistical model selection, ranging from the fact that all model selection methods show particular weaknesses in certain situations, to differences in the goals of computational modeling.

37. •• Nussenbaum K, Hartley CA: **Reinforcement learning across development: what insights can we draw from a decade of research?** *Dev Cogn Neurosci* 2019, **40**:100733 http://dx.doi.org/10.1016/j.dcn.2019.100733

The authors provide an excellent review of the developmental computational modeling research of the past decade, with a focus on describing the discrepancies that have emerged, and offering initial explanations.

38. Hauser TU, Will G-J, Dubois M, Dolan RJ: **Annual research review: developmental computational psychiatry**. *J Child Psychol Psychiatry* 2019, **60**:412-426 http://dx.doi.org/10.1111/jcpp.12964.

39. Huys QJM, Maia TV, Frank MJ: **Computational psychiatry as a bridge from neuroscience to clinical applications**. *Nat Neurosci* 2016, **19**:404-413 http://dx.doi.org/10.1038/nn.4238.

40. Teller DY: **Linking propositions**. *Vision Res* 1984, **24**:1233-1246 http://dx.doi.org/10.1016/0042-6989(84)90178-0.

41. •• Brown VM, Chen J, Gillan CM, Price RB: **Improving the reliability of computational analyses: model-based planning and its relationship with compulsivity**. *Biol Psychiatry: Cogn Neurosci Neuroimaging* 2020, **5**:601-609 http://dx.doi.org/10.1016/j.bpsc.2019.12.019

A careful methodological investigation assessing the test–retest and split-half reliability of the crucial mixing parameter of a famous RL model, revealing large variability based on the chosen analysis approach.

42. Daniel R, Radulescu A, Niv Y: **Intact reinforcement learning but impaired attentional control during multidimensional probabilistic learning in older adults**. *J Neurosci* 2020, **40**:1084-1096 http://dx.doi.org/10.1523/JNEUROSCI.0254-19.2019.

43. Kaiser RH, Treadway MT, Wooten DW, Kumar P, Goer F, Murray L, Beltzer M, Pechtel P, Whitton A, Cohen AL, Alpert NM, El Fakhri G, Normandin MD, Pizzagalli DA: **Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation**. *Cereb Cortex* 2018, **28**:4281-4290 http://dx.doi.org/10.1093/cercor/bhx281.

44. Javadi AH, Schmidt DHK, Smolka MN: **Adolescents adapt more slowly than adults to varying reward contingencies**. *J Cogn Neurosci* 2014, **26**:2670-2681 http://dx.doi.org/10.1162/jocn_a_00677.

45. Gershman SJ: **Empirical priors for reinforcement learning models**. *J Math Psychol* 2016, **71**:1-6 http://dx.doi.org/10.1016/j.jmp.2016.01.006.

46. Kool W, Cushman FA, Gershman SJ: **When does model-based control pay off?** *PLOS Comput Biol* 2016, **12**:e1005090 http://dx.doi.org/10.1371/journal.pcbi.1005090.

47. Harada T: **Learning from success or failure? — Positivity biases revisited**. *Front Psychol* 2020, **11** http://dx.doi.org/10.3389/fpsyg.2020.01627.

48. Gerraty RT, Davidow JY, Foerde K, Galvan A, Bassett DS, Shohamy D: **Dynamic flexibility in striatal-cortical circuits supports reinforcement learning**. *J Neurosci* 2018, **38**:2442-2453 http://dx.doi.org/10.1523/JNEUROSCI.2084-17.2018.

49. Watabe-Uchida M, Eshel N, Uchida N: **Neural circuitry of reward prediction error**. *Annu Rev Neurosci* 2017, **40**:373-394 http://dx.doi.org/10.1146/annurev-neuro-072116-031109.

50. Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M: **A distributional code for value in dopamine-based reinforcement learning**. *Nature* 2020, **577**:671-675 http://dx.doi.org/10.1038/s41586-019-1924-6.

51. Collins AGE, Frank MJ: **Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive**. *Psychol Rev* 2014, **121**:337-366 http://dx.doi.org/10.1037/a0037015.

52. Tai L-H, Lee AM, Benavidez N, Bonci A, Wilbrecht L: **Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value**. *Nat Neurosci* 2012, **15**:1281-1289 http://dx.doi.org/10.1038/nn.3188.

53. Cox J, Witten IB: **Striatal circuits for reward learning and decision-making**. *Nat Rev Neurosci* 2019, **20**:482-494 http://dx.doi.org/10.1038/s41583-019-0189-2.

54. Rupprechter S, Romaniuk L, Series P, Hirose Y, Hawkins E, Sandu A-L, Waiter GD, McNeil CJ, Shen X, Harris MA, Campbell A, Porteous D, Macfarlane JA, Lawrie SM, Murray AD, Delgado MR, McIntosh AM, Whalley HC, Steele JD: **Blunted medial prefrontal**

cortico-limbic reward-related effective connectivity and depression. *Brain* 2020, **143**:1946-1956 http://dx.doi.org/10.1093/brain/awaa106.

55. van den Bos W, Bruckner R, Nassar MR, Mata R, Eppinger B: **Computational neuroscience across the lifespan: promises and pitfalls**. *Dev Cogn Neurosci* 2017 http://dx.doi.org/10.1016/j.dcn.2017.09.008.

56. Bolenz F, Reiter AMF, Eppinger B: **Developmental changes in learning: computational mechanisms and social influences**. *Front Psychol* 2017, **8** http://dx.doi.org/10.3389/fpsyg.2017.02048.

57. Yaple ZA, Yu R: **Fractionating adaptive learning: a meta-analysis of the reversal learning paradigm**. *Neurosci Biobehav Rev* 2019, **102**:85-94 http://dx.doi.org/10.1016/j.neubiorev.2019.04.006.

58. O'Doherty JP, Lee SW, McNamee D: **The structure of reinforcement-learning mechanisms in the human brain**. *Curr Opin Behav Sci* 2015, **1**:94-100 http://dx.doi.org/10.1016/j.cobeha.2014.10.004.

59. Garrison J, Erdeniz B, Done J: **Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies**. *Neurosci Biobehav Rev* 2013, **37**:1297-1310 http://dx.doi.org/10.1016/j.neubiorev.2013.03.023.

60. Lee D, Seo H, Jung MW: **Neural basis of reinforcement learning and decision making**. *Annu Rev Neurosci* 2012, **35**:287-308 http://dx.doi.org/10.1146/annurev-neuro-062111-150512.

61. Yarkoni T: **The generalizability crisis**. *Behav Brain Sci* 2020
•• http://dx.doi.org/10.1017/S0140525X20001685
A thoughtful and thorough critique on the omnipresent overinterpretation of specific findings as general rules and on the lack of robustness tests that would alleviate this issue in psychological science.

62. Liu X, Hairston J, Schrier M, Fan J: **Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies**. *Neurosci Biobehav Rev* 2011, **35**:1219-1236 http://dx.doi.org/10.1016/j.neubiorev.2010.12.012.

63. Davidow J, Foerde K, Galvan A, Shohamy D: **An upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence**. *Neuron* 2016, **92**:93-99 http://dx.doi.org/10.1016/j.neuron.2016.08.031.

64. Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS: **Learning the value of information in an uncertain world**. *Nat Neurosci* 2007, **10**:1214-1221 http://dx.doi.org/10.1038/nn1954.

65. Lefebvre G, Lebreton M, Meyniel F, Bourgeois-Gironde S, Palminteri S: **Behavioural and neural characterization of optimistic reinforcement learning**. *Nat Hum Behav* 2017, **1**:0067 http://dx.doi.org/10.1038/s41562-017-0067.

66. Daw N, Gershman S, Seymour B, Dayan P, Dolan R: **Model-based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215 http://dx.doi.org/10.1016/j.neuron.2011.02.027.

67. Katahira K: **The statistical structures of reinforcement learning with asymmetric value updates**. *J Math Psychol* 2018, **87**:31-45 http://dx.doi.org/10.1016/j.jmp.2018.09.002.

68. Sugawara M, Katahira K: **Dissociation between asymmetric**
• **value updating and perseverance in human reinforcement learning**. *Sci Rep* 2021, **11**:3574 http://dx.doi.org/10.1038/s41598-020-80593-7
The detailed model simulation study reveals that behavior caused by perseverance can falsely be attributed to an asymmetry between learning rates from positive and negative outcomes, and vice versa; only a hybrid model combining perseverance and asymmetric learning rates disentangles both factors.

69. Xia L, Master S, Eckstein M, Wilbrecht L, Collins AGE: **Learning under uncertainty changes during adolescence**. *Proceedings of the Cognitive Science Society* 2020.

70. Decker JH, Lourenco FS, Doll BB, Hartley CA: **Experiential reward learning outweighs instruction prior to adulthood**. *Cogn Affect Behav Neurosci* 2015, **15**:310-320 http://dx.doi.org/10.3758/s13415-014-0332-5.

71. Palminteri S, Kilford EJ, Coricelli G, Blakemore S-J: **The computational development of reinforcement learning during adolescence**. *PLoS Comput Biol* 2016, **12** http://dx.doi.org/10.1371/journal.pcbi.1004953.

72. Deserno L, Boehme R, Heinz A, Schlagenhauf F: **Reinforcement learning and dopamine in schizophrenia: dimensions of symptoms or specific features of a disease group?** *Front Psychiatry* 2013, **4** http://dx.doi.org/10.3389/fpsyt.2013.00172.

73. Ahn W-Y, Busemeyer JR: **Challenges and promises for translating computational tools into clinical practice**. *Curr Opin Behav Sci* 2016, **11**:1-7 http://dx.doi.org/10.1016/j.cobeha.2016.02.001.

74. Blakemore S-J, Robbins TW: **Decision-making in the adolescent brain**. *Nat Neurosci* 2012, **15**:1184-1191 http://dx.doi.org/10.1038/nn.3177.

75. DePasque S, Galván A: **Frontostriatal development and probabilistic reinforcement learning during adolescence**. *Neurobiol Learn Mem* 2017, **143**:1-7 http://dx.doi.org/10.1016/j.nlm.2017.04.009.

76. Eckstein MK, Master SL, Xia L, Dahl RE, Wilbrecht L, Collins AGE:
•• **Learning rates are not all the same: the interpretation of computational model parameters depends on the context**. *bioRxiv* 2021 http://dx.doi.org/10.1101/2021.05.28.446162
An empirical investigation of the generalizability and interpretability of RL model parameters, employing a within-participants design to assess correlations and shared variance between the same parameters across tasks.

77. Groman SM, Keistler C, Keip AJ, Hammarlund E, DiLeone RJ, Pittenger C, Lee D, Taylor JR: **Orbitofrontal circuits control multiple reinforcement-learning processes**. *Neuron* 2019, **103**:734-746.e3 http://dx.doi.org/10.1016/j.neuron.2019.05.042.

78. Starkweather CK, Gershman SJ, Uchida N: **The medial prefrontal cortex shapes dopamine reward prediction errors under state uncertainty**. *Neuron* 2018, **98**:616-629.e6 http://dx.doi.org/10.1016/j.neuron.2018.03.036.

79. Gershman SJ, Uchida N: **Believing in dopamine**. *Nat Rev Neurosci* 2019, **20**:703-714 http://dx.doi.org/10.1038/s41583-019-0220-7.

80. Frömer R, Dean Wolf CK, Shenhav A: **Goal congruency**
• **dominates reward value in accounting for behavioral and neural correlates of value-based decision-making**. *Nat Commun* 2019, **10**:4926 http://dx.doi.org/10.1038/s41467-019-12931-x
A carefully designed, multi-experiment study that shows that the brain's 'value network' might instead encode the congruency of choice options with participants' goals, contradicting a large body of previous RL research, and highlighting the major impact of a common experimental confound.

81. van den Bos W, Hertwig R: **Adolescents display distinctive tolerance to ambiguity and to uncertainty during risky decision making**. *Sci Rep* 2017, **7**:40962 http://dx.doi.org/10.1038/srep40962.

82. Sendhilnathan N, Semework M, Goldberg ME, Ipata AE: **Neural correlates of reinforcement learning in mid-lateral cerebellum**. *Neuron* 2020, **106**:188-198.e5 http://dx.doi.org/10.1016/j.neuron.2019.12.032.

83. McDougle SD, Collins AGE: **Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning**. *Psychonom Bull Rev* 2021, **28**:20-39 http://dx.doi.org/10.3758/s13423-020-01774-z.

84. Konovalov A, Krajbich I: **Neurocomputational dynamics of sequence learning**. *Neuron* 2018, **98**:1282-1293.e4 http://dx.doi.org/10.1016/j.neuron.2018.05.013.

85. Kalashnikov D, Irpan A, Pastor P, Ibarz J, Herzog A, Jang E, Quillen D, Holly E, Kalakrishnan M, Vanhoucke V, Levine S: *QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation*. 2018arXiv:1806.10293 [cs, stat].

86. Bakkour A, Palombo DJ, Zylberberg A, Kang YH, Reid A, Verfaellie M, Shadlen MN, Shohamy D: **The hippocampus supports deliberation during value-based decisions**. *eLife* 2019, **8**:e46080 http://dx.doi.org/10.7554/eLife.46080.

87. Momennejad I, Russek EM, Cheong JH, Botvinick M, Daw ND, Gershman SJ: **The successor representation in human reinforcement learning**. *Nat Hum Behav* 2017, **1**:680-692 http://dx.doi.org/10.1038/s41562-017-0180-8.