

# Visual novelty, curiosity, and intrinsic reward in machine learning and the brain

Andrew Jaegle, Vahid Mehrpour and Nicole Rust



A strong preference for novelty emerges in infancy and is prevalent across the animal kingdom. When incorporated into reinforcement-based machine learning algorithms, visual novelty can act as an intrinsic reward signal that vastly increases the efficiency of exploration and expedites learning, particularly in situations where external rewards are difficult to obtain. Here we review parallels between recent developments in novelty-driven machine learning algorithms and our understanding of how visual novelty is computed and signaled in the primate brain. We propose that in the visual system, novelty representations are not configured with the principal goal of detecting novel objects, but rather with the broader goal of flexibly generalizing novelty information across different states in the service of driving novelty-based learning.

## Address

Department of Psychology, University of Pennsylvania, United States

Corresponding author: Rust, Nicole ([nrust@sas.upenn.edu](mailto:nrust@sas.upenn.edu))

**Current Opinion in Neurobiology** 2019, **58**:167–174

This review comes from a themed issue on **Computational neuroscience**

Edited by **Máté Lengyel** and **Brent Doiron**

<https://doi.org/10.1016/j.conb.2019.08.004>

0959-4388/© 2019 Elsevier Ltd. All rights reserved.

## Introduction

What signals does an animal or other agent need to learn to generate good behavior? Reinforcement learning (RL) provides one answer to this question: to learn to act, an agent needs access to a reward signal, and it needs to estimate which states and actions cause that signal to be large [reviewed by Ref. 1]. Typically, RL assumes that the reward signal comes in the form of an external reward, such as a satisfying food. RL has proven highly effective for describing reward-based learning in neural systems [2], and it has led to breakthroughs in machine learning [3,4]. However, classic implementations of RL are most effective when reward signals are easily and frequently obtained [1], and this is typically not the case in real environments. What type of signal can be used to learn in settings where rewards are rare?

Curiosity in the form of novelty seeking is one candidate for such a signal [5]. Human infants exhibit strong preferences for stimuli and situations unlike those encountered before, even when rewards are not otherwise associated with these stimuli [6]. Analogous novelty preferences have been observed in over 100 species, ranging from reptiles to monkeys [7,8]. Stimulus novelty is an important driver of activity in the midbrain dopaminergic system, which plays a well-known role in signaling the rewarding value of stimuli and consequently has been implicated in accounts of how the brain performs RL-like computations [59,60]. Recent work in machine learning demonstrates how novelty can be incorporated into RL algorithms to successfully drive the mastery of complex tasks. In parallel, recent developments in neuroscience have provided important insights into how visual novelty is computed and signaled in the brain. In this review, we synthesize recent work on the computation, representation, and uses of visual novelty in both machine learning and neuroscience. By drawing these connections, we hope to encourage more rapid progress in both fields.

## Visual novelty and intrinsic motivation for machine-based RL

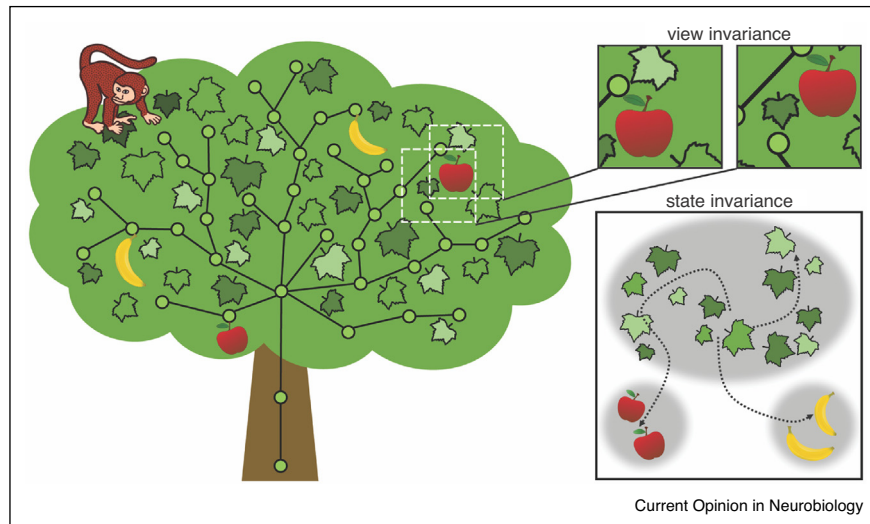
Consider a monkey foraging for fruit hidden in the dense brush of a tree, where different branches can be envisioned as pathways that lead to different ‘states’ where the fruit might be (Figure 1). How might the monkey decide which states are valuable? When fruit is sparse, the monkey is unlikely to find rewards by exploring branches purely at random, as only a few states contain fruit, and it will typically miss these states, often revisiting fruitless parts of the tree. A more effective strategy is to explore branches that are different than those tried in the past. This can be achieved by making the act of exploring novel things itself a rewarding experience for the monkey, and this is precisely what novelty-driven RL algorithms seek to do.

Classic RL algorithms learn what outcomes lead to what rewards by assigning values to different states. Novelty-based RL methods seek to incorporate the novelty of a state into its value by augmenting the classic equation for computing value, the Bellman equation [9], with a third term that reflects the state’s novelty:

$$V(s) = \max_{a \in A} [R(s, a) + \gamma \mathbb{E}_p[V(s')]] + \beta N(s)$$

where  $V(s)$  is the value of the current state,  $s$ . The first term of this equation,  $R(s, a)$ , is an estimate of the reward that the agent will get by taking one action ( $a$ ) from the

Figure 1



Novelty can drive exploration in environments with sparse external rewards.

An illustration of the benefits of exploration: a monkey is trying to find a piece of fruit (a reward) in a large, densely foliated tree with many branches. Typically, the monkey must make many choices and explore many 'states' before it receives a single reward. If the monkey finds novel states rewarding, then it will be encouraged to explore the tree, and it can discover rewarding states that it would otherwise miss. *View invariance*: different views of the same state (e.g. the apple) can correspond to different images. To effectively drive RL, a system must map images onto their corresponding states. *State invariance*: different states can share features that are indicative of their novelty, for example, reflecting the fact that fruits are usually large and are rarely green while leaves are often small and can take on many shades of green. A system that can exploit the features shared by different states can drive the monkey to explore states with novel features (e.g. objects with a new size or shade).

set of possible actions ( $A$ ). The second term,  $\mathbb{E}_p[V(s')]$ , is the anticipated value of the next state,  $s'$ , taking into account the expected dynamics of the environment (described by a transition function  $p(s'|s, a)$ ). Here  $\gamma$  is a discount factor used to balance the reward from the current state with the expected value of future states. Together, these first two terms capture the original Bellman formulation, which does not incorporate novelty considerations. The final term,  $N(s)$ , was incorporated later to assign value based on the state's novelty, where  $\beta$  is a parameter that balances reward seeking with novelty seeking [10].

In simple environments, novelty can be computed by keeping track of all states that have been visited, and states are rated as more novel if they have been visited less frequently. Algorithms that employ such an approach are often called *count-based strategies*. In early proposals for novelty-driven RL, count-based strategies were demonstrated to lead to good exploration behavior in simple RL settings such as bandit problems and simple Markov decision processes [11–13].

Contemporary RL approaches typically address the problem of training an agent to produce a controller or motor command given the pixel pattern on a video display, in settings such as Atari video games [14] or navigation through complex 3D scenes (Beattie *et al.*, arXiv: 1612.03801). To effectively estimate and

use novelty in these settings, a system must overcome the high-dimensionality inherent to these problems, a challenge that can be parsed into two conceptually distinct invariance problems. The first is *view invariance*: because the same underlying state can be viewed from many different vantage points, a system has to sort out the mapping from a diversity of images onto their corresponding states to estimate the novelty of the current visual input (Figure 1). However, this alone is not sufficient to explore settings with a large number of states, where exhaustive exploration of unique states remains intractable. Rather, a system must also be capable of *state invariance*: capturing the groupings of different states based on their shared characteristics, such as the fact that ripe fruits are rarely green (Figure 1). In contemporary novelty-driven RL algorithms, state-invariant estimates of novelty are a crucial component of guiding exploration toward states that are most unlike those previously seen and efficiently finding previously undiscovered rewards.

Recent novelty-driven RL methods address both invariance problems by employing some form of a model that transforms images into novelty estimates. Most methods train a deep neural network (DNN) for this goal, as DNNs have proven successful at learning invariant and generalizable representations in a broad range of settings (for reviews, see Refs. [15,16]). Recent proposals differ in terms of the details by which these models are trained and how the resulting measures of image similarity are converted into

estimates of novelty, as described below. Once estimated, nearly all proposals incorporate novelty into measures of value in the same way (i.e. by incorporating novelty into the Bellman equation), as described above.

One class of recent proposals for novelty-driven RL addressed the view and state invariance challenges by incorporating a method for computing similarity between images. Some of these proposals extend classic count-based proposals by first mapping similar images to the same bin (e.g. using a hash function [17<sup>•</sup>] or by clustering [18]) before counting. A complementary proposal [19<sup>•</sup>] estimated image similarity by training a DNN to estimate the distance in time between pairs of images, motivated by the observation that temporal continuity can capture invariance [20]. Instead of counting how many times states had been visited, this method estimated an image's novelty by comparing it to familiar images held in a memory buffer.

Another class of proposals for novelty-driven RL used 'pseudocounts' to estimate novelty [21<sup>••</sup>,22,23]. In contrast to explicit count-based proposals, pseudocount methods address the invariance problems using a model (e.g. a DNN) trained to estimate the probability of images (i.e. the probability that images will be seen by an agent over its lifetime). In these models, image probability is estimated using a model which is trained to generate images by predicting the probability distribution over the pixel values of an image and then sampling from this distribution (a so-called 'generative model'). By maximizing the probability of generating images from the training set, models of this kind learn to evaluate the probability of any image, including those not encountered during training. Maximum likelihood training of this kind is a type of unsupervised learning, as it does not require images to be labeled but instead relies on image generation to learn a model of images.

When trained to maximize the probability of images seen, images that have been seen before will be scored as more likely than images that have never been seen, and after training, such a model can be used to estimate the probability of an input image (i.e. its likelihood under the trained model). Pseudocounts are computed based on how the probability of an image changes between the  $N$ th and  $(N + 1)$ th times it is viewed, where  $N$  is often zero [21<sup>••</sup>]. To understand why the change in the probability reflects the count, consider the empirical probability of an image: when the image has been seen  $k$  times among the first  $n$  images, its empirical probability is  $p_1 = \frac{k}{n}$ . When it is seen again, its probability is  $p_2 = \frac{k+1}{n+1}$ . We can rewrite these two equations to give an expression for the count:  $k = \frac{p_1(1-p_2)}{(p_2-p_1)}$ . This expression can be computed using only the probability of an image before and after it is viewed. Because the pseudocount method uses the continuous

probability estimated by a model, and not a discrete count, it is well suited to scenarios where the dimensionality is high and hence nearly all counts are zero. Continuous probabilities naturally allow for (1) the representation of arbitrarily small probabilities (the probability of seeing any given image is small, as any specific image will typically be experienced very infrequently) and (2) the representation of small and variable changes in probability (the change in probability of an image over the course of experience depends not only on how many times that image has been seen, but how frequently similar or dissimilar images have been seen). Pseudocount methods rely on the change in response to an image after exposure to it, and in this respect they bare some similarity to the phenomenon of *repetition suppression* in the visual system (see below for a discussion of the role of repetition suppression in novelty computation in the brain). Several recent papers have achieved promising results using pseudocounts to drive exploration on difficult RL tasks [21<sup>••</sup>,22,23].

How does novelty seeking relate to other types of intrinsic motivation? In the case of pseudocount estimates, an important theoretical link has been established between estimates of novelty and the amount of *information gain* [24] that follows from observing an image [21<sup>••</sup>]. Intuitively, this is because the novelty of an image reflects how much it differs from what we expect to see based on what we have seen in the past. Early work on intrinsic motivation established the link between curiosity and measures of image informativeness, including information gain [25,26]. Information maximization and related objectives have also been widely used to account for sensory exploration in animal behavior [61]. While the information gained by observing an image cannot be computed directly, several methods have been proposed to approximate it to drive exploration [27,28]. A number of recent papers have also proposed other, related signals to drive exploration, including methods that estimate image informativeness by how variable or unreliable the response to the image is [29], how well the response of a target model can be predicted [30], and by how difficult it is to predict what will follow an image in time [31–33]. Unlike methods for novelty-driven RL, these methods do not estimate novelty explicitly, but they share the goal of driving exploration by estimating the informativeness of images. Other methods also encourage appropriate levels of exploration by taking into account the uncertainty of state value estimates (including classical methods using Gittins indices [34,35] and Thompson sampling [36,37]), but the relationship of this work to the novelty signals we review here is less clear.

In sum, machine learning algorithms that incorporate novelty as an intrinsic reward have proven effective at driving RL, particularly in environments where rewards are sparse. To compute novelty, contemporary algorithms are designed to generalize across different views of the

same state and to share features across different states (Figure 1). These algorithms estimate the novelty of the current visual input using a variety of techniques, but we can broadly distinguish between those that employ explicit count-based methods of novelty and those that estimate novelty using pseudocount procedures. Direct theoretical links have been established between pseudocounts and the amount of information gained by viewing an image. All methods share the goal of driving exploration toward states that have not yet been adequately explored. In the next section, we focus on recent progress toward understanding where and how correlates of these novelty signals may be computed and signaled in the primate brain.

### Computing and signaling novelty in the primate brain

Our ability to detect visual novelty (or equivalently remember whether we have seen an image) is quite remarkable – for example, we can view thousands of photographs, each only once and each for a few seconds, and then distinguish with high accuracy the specific images that we have already seen from those that remain novel to us [38]. Even after viewing as many as 10 000 distinct photographs, our rates of remembering do not saturate, suggesting that this type of single-exposure visual memory has an exceedingly large capacity [38,39]. The most remarkable aspects of our ability to detect visual novelty are thought to be mediated by our ‘familiarity’ memory system. One effective illustration of familiarity is the experience that we all occasionally have of seeing someone and remembering that we know them but not being able to recall any details about them, at least for a few moments, and this ‘sense of remembering absent details’ is precisely what the familiarity memory system supports. In contrast to recollection-based memories (e.g. of the details about that person, such as their name), which are thought to be largely mediated by the hippocampus, familiarity is thought to be mediated by another brain area in the medial temporal lobe, perirhinal cortex, as well its input from the part of the visual system involved in signaling object and scene information, inferotemporal cortex (IT) [reviewed by Refs. 40,41].

How do these brain areas signal visual novelty? Novelty is thought to be signaled in IT and perirhinal cortex via an adaptation-like change in firing rate in response to familiar as compared to novel stimuli, a phenomenon referred to as *repetition suppression* [42–45,46\*\*]. Consistent with the signatures needed to account for the vast capacity of human single-exposure visual memory behavior, IT and perirhinal firing rate reductions with familiarity are selective for images, even after viewing large numbers of them, and these response decrements last for several minutes to hours following the single viewing of an image [43,44,46\*\*]. The representations of novelty and familiarity are best understood in IT, where these putative visual novelty signals are

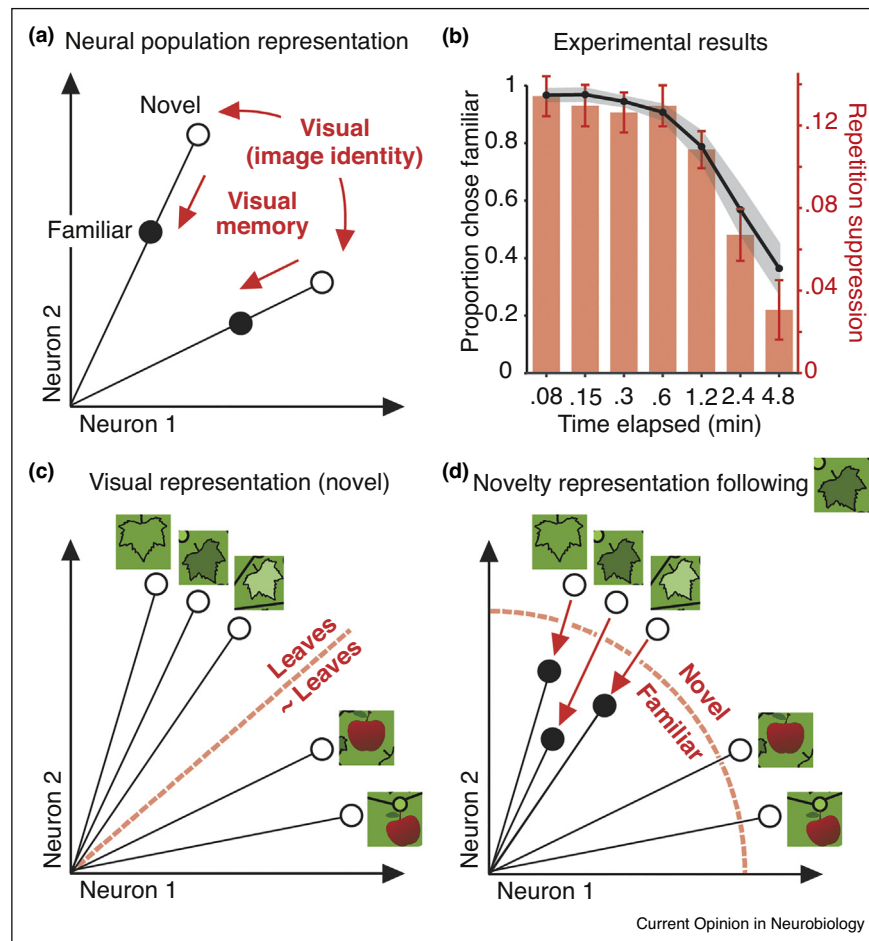
mixed with signals reflecting visual identity, both within the responses of individual neurons and across the IT population. That is, visual identities of images and their content are thought to be reflected as distinct patterns of spikes across the IT population, and this translates into a population representation in which visual information about the currently viewed scene is reflected by a population’s vector angle [Figure 2; reviewed by Ref. 47]. In contrast, novelty is thought to be signaled by overall firing rates or equivalently the length of the population response vector, where vectors are longer for novel images and become shorter as they become familiar [46\*\*]. Novelty and familiarity modulations are thought to be approximately multiplicative [48,49], which translates to a type of novelty coding for memory that maintains identity vector angle position, thereby preventing the interference of identity and novelty representations [46\*\*]. Repetition suppression magnitudes are also continuous and depend on factors such as the time since an image has been viewed, the duration of the viewing, and the number of repeated viewings [reviewed by Ref. 50]. This encoding scheme can account for behavior on a familiarity task with a decoder that maps IT neural responses to behavior via a simple positively weighted linear read-out that sums the responses of the top 50% of most memory sensitive IT neurons ([46\*\*], Figure 2b).

How is novelty computed by the brain? In the framework described in Figure 2, this amounts to understanding the origin of IT repetition suppression. Repetition suppression is found at all stages of visual processing from the retina to IT and perirhinal cortex, and it strengthens in its magnitude as well as the duration over which it lasts across the visual cortical hierarchy [51]. Consequently, a hierarchical cascade of feed-forward, adaptation-like mechanisms clearly contribute to IT repetition suppression [50]. There are also indications that IT repetition suppression may arise from changes in synaptic weights between recurrently connected units within IT [50,52] and/or feed-back mechanisms from higher brain areas (such as perirhinal cortex) [53,54], although the latter assertion has been the focus of some debate [reviewed by Ref. 50].

### Insights from comparisons between brain-based and machine-based approaches to novelty

Machine learning and neuroscience arrive at novelty considerations from complementary perspectives: the former solves challenging engineering problems whereas the latter discovers the principles by which the brain operates. Despite their different objectives, they have much to learn from one another. For example, insofar as the brain is a machine that can perform these tasks, engineering insight can be gained from an understanding of how it operates. Conversely, engineering considerations can provide insights into the rationale behind

Figure 2



The representation of visual novelty in IT.

**(a)** Shown are the hypothetical population responses to two images, each presented as both novel and familiar, plotted as the spike count response of neuron 1 versus neuron 2. In this scenario, visual identity (e.g. image or object identity) is reflected by the population response pattern, or equivalently, the direction of each population response vector. In contrast, novelty is reflected by the population vector length, where images with longer population vectors are more novel, and novelty can be extracted with a simple, positively weighted decoder. **(b)** The alignment of behavioral reports of remembering images and IT repetition suppression, reported by Ref. [46\*\*], in an experiment configured such that memories of images last minutes. The behavioral 'forgetting function' (black) depicts the proportion of trials that rhesus monkeys remembered images as a function of the time since seeing those images as novel. This was determined to be well-aligned with the magnitude of IT repetition suppression in the top 50% of most sensitive IT memory neurons, where repetition suppression is quantified as the mean proportional reduction in the response to familiar images relative to the responses when those images were novel. **(c)** Visual representations in IT (depicted here for novel images) are grouped by object identity and category similarity and formatted such that identity and category (e.g. leaves) can be extracted with a simple linear decoder. **(d)** Repetition suppression is strongest for repeated presentations of the same image but also has partial transfer across images that are similar.

specific brain implementations. From machine learning, we learn that novelty is a widely useful tool for finding solutions to problems where environmental rewards are sparse. For a novelty signal to be useful in real settings, it must generalize to new images and new scenes. From neuroscience, we learn that visual novelty is computed using a specialized memory system that is tightly integrated with the system for visual recognition. The computation of this system does not capture all of the details required for full recollection, but instead generalizes smoothly between images and over time. Here we

highlight two insights that emerge from the comparison between brain-based and machine-based approaches to novelty that we find particularly striking.

**The signals for novelty are continuous and are multiplexed with identity information in the brain as well as in a subset of machine-based approaches**

As described above, the brain does not parcellate the signaling of visual identity and visual novelty into distinct modules, but rather multiplexes identity and novelty representations at the highest stages of the visual system

(Figure 2a). Unsurprisingly, brain-based measures of novelty are continuous, as opposed to binary or discrete. While continuous or graded signals are not needed to represent novelty in simple state spaces, they are needed to represent novelty in light of the generalization challenges presented by real images. The brain's strategy of multiplexing identity and novelty signals in a high-dimensional continuous way appear to be analogous to techniques (like pseudocounts) that compute continuous measures of novelty by comparing continuous probabilities across repeated image exposures using recognition architectures. The implications of this connection are still under-explored and remain an important direction for future work. As described above, this type of encoding scheme is highly advantageous for novelty-driven learning when the dimensionality is large and nearly everything is novel.

### The brain supports representations of visual novelty that are tolerant rather than invariant, thereby making them flexible for different tasks and environments

To summarize one insight from above, novelty-driven RL needs to estimate view and state invariant measures of similarity between the current visual input and previously viewed images (Figure 1). In the brain, novelty information (in the form of repetition suppression) is multiplexed with visual information in IT, the same brain area implicated in the computation of object identity 'invariant' to view [reviewed by Ref. 47]. While it is thus natural to conjecture that IT novelty representations are configured with the principal goal of detecting the appearance of novel objects, we suspect that this is only part of the story. Rather, as demonstrated above, the 'states' that need to be explored to solve real-world learning problems (and analogously, Atari video games) are not typically defined by the objects contained therein, but rather by something more akin to scenes, which delimit and ground how objects are perceived and used [see also Ref. 55]. There are several indications that IT novelty representations are configured with the broader goal of flexibly generalizing novelty information across different states. *First*, while not broadly appreciated, the responses of individual IT neurons are not themselves 'invariant' but rather tolerant, meaning they remain sensitive to changes in parameters such as the position, size, and background context with which objects appear. Invariance emerges across the population as a consequence of the linearly separable format of IT representations, where object identity can be extracted with a simple linear decoding scheme (Figure 2c, [reviewed by Ref. 47]). Moreover, IT reflects explicit (i.e. linearly separable) information about scene details (such as an object's position) to a higher degree than any earlier stage of the form processing pathway [56\*]. *Second*, our behavioral reports of novelty contain rich information about the context and configurations that objects appear in Ref. [39], suggesting that novelty-based computations are not object-invariant, but

rather incorporate considerable detail about scenes. *Third*, while novelty representations in IT are generally very selective for image identity [57], after viewing an image, repetition suppression does in fact transfer to other, similar images ([58], Figure 2d). This includes transfer across images whose similarity is conferred by their shape, as well as other variables such as an object's spatial position. Taken together, these results suggest that graded IT novelty representations reflect information about the images that have been encountered but also support novelty generalization across images that are similar in virtue of either their view or state (Figure 1). We speculate that the representation of identity and novelty in IT (as illustrated in Figure 2a, c–d) supports adaptation to new environments by enabling downstream behavioral decoding to reflect the relationship of new images to those encountered in past experience.

### Conclusions

The major successes in AI have largely come in settings with frequent, identifiable rewards, but animals live in complex environments that do not provide ubiquitous, easily identifiable rewards. Yet human infants, and other developing animals, are able to reliably produce complex behaviors with only a small amount of experience. The success of novelty-based approaches in overcoming the limitations of current approaches to learning suggests that novelty computation may play a crucial role in how animals explore and learn. Signals reflecting visual novelty are found in the primate brain in higher visual areas such as IT, where they are reflected as continuous measures of novelty and multiplexed with representations of visual identity. Taken together, recent studies suggest that IT novelty representations can generalize novelty across different views and states in the manner required to drive novelty-based exploration and learning.

### Funding

This work was supported by the National Eye Institute of the US National Institutes of Health (award R01EY020851), the National Science Foundation (award 1265480), and the Simons Foundation (through an award from the Simons Collaboration on the Global Brain, award 543033).

### Conflict of interest statement

Nothing declared.

### 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. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. MIT Press; 2017.
  2. Lee D, Seo H, Jung MW: **Neural basis of reinforcement learning and decision making**. *Annu Rev Neurosci* 2012, **35**:287–308.

3. Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A: **Mastering the game of Go without human knowledge.** *Nature* 2017, **550**:354.
  4. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G: **Human-level control through deep reinforcement learning.** *Nature* 2015, **518**:529.
  5. Gottlieb J, Oudeyer P-Y: **Towards a neuroscience of active sampling and curiosity.** *Nat Rev Neurosci* 2018:1.
  6. Reynolds GD: **Infant visual attention and object recognition.** *Behav Brain Res* 2015, **285**:34-43.
  7. Glickman SE, Sroges RW: **Curiosity in zoo animals.** *Behaviour* 1966:151-188.
  8. Hall BA, Melfi V, Burns A, McGill DM, Doyle RE: **Curious creatures: a multi-taxa investigation of responses to novelty in a zoo environment.** *PeerJ* 2018, **6**:e4454.
  9. Bellman R: **The theory of dynamic programming.** *Bull Am Math Soc* 1954, **60**:503-515.
  10. Strehl AL, Littman ML: **An analysis of model-based interval estimation for Markov decision processes.** *J Comput Syst Sci* 2008, **74**:1309-1331.
  11. Kolter JZ, Ng AY: **Near-Bayesian exploration in polynomial time.** *International Conference on Machine Learning (ICML)*. 2009:513-520.
  12. Lai TL, Robbins H: **Asymptotically efficient adaptive allocation rules.** *Adv Appl Math* 1985, **6**:4-22.
  13. Strehl AL, Littman ML: **A theoretical analysis of model-based interval estimation.** *International Conference on Machine Learning (ICML)*. 2005:856-863.
  14. Bellemare MG, Naddaf Y, Veness J, Bowling M: **The arcade learning environment: an evaluation platform for general agents.** *J Artif Intell Res* 2013, **47**:253-279.
  15. Goodfellow I, Bengio Y, Courville A, Bengio Y: *Deep Learning*. Cambridge: MIT Press; 2016.
  16. LeCun Y, Bengio Y, Hinton G: **Deep learning.** *Nature* 2015, **521**:436.
  17. Tang H, Houthoofd R, Foote D, Stooke A, Chen OX, Duan Y, Schulman J, DeTurck F, Abbeel P: **#Exploration: a study of count-based exploration for deep reinforcement learning.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2017:2753-2762.
- This paper addressed the view and state invariance challenges by adapting count-based methods to image-based RL using hash functions that map images to a relatively small number of bins.
18. Abel D, Agarwal A, Diaz F, Krishnamurthy A, Schapire R: **Exploratory gradient boosting for reinforcement learning in complex domains.** *International Conference on Machine Learning (ICML) Workshop on Abstraction in Reinforcement Learning*. 2016.
  19. Savinov N, Raichuk A, Marinier R, Vincent D, Pollefeys M, Lillicrap T, Gelly S: **Episodic curiosity through reachability.** *International Conference on Learning Representations (ICLR)*. 2019.
- This paper computed an intrinsic reward signal that closely resembles a novelty computation using two interesting model components: a learned similarity measure and a memory buffer.
20. Anselmi F, Leibo JZ, Rosasco L, Mutch J, Tacchetti A, Poggio T: **Unsupervised learning of invariant representations.** *Theor Comput Sci* 2016, **633**:112-121.
  21. Bellemare M, Srinivasan S, Ostrovski G, Schaul T, Saxton D, Munos R: **Unifying count-based exploration and intrinsic motivation.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2016:1471-1479.
- This paper introduced the notion of pseudocounts, which generalize count-based methods for novelty estimation and can be applied to RL problems with high-dimensional, continuous states such as images. The authors formally demonstrated the relationship between pseudocounts and information gain, suggesting that pseudocounts may lead to near-optimal exploration behavior.
22. Martin J, Sasikumar SN, Everitt T, Hutter M: **Count-based exploration in feature space for reinforcement learning.** *International Joint Conference on Artificial Intelligence (IJCAI)*. 2017.
  23. Ostrovski G, Bellemare MG, Avd Oord, Munos R: **Count-based exploration with neural density models.** *International Conference on Machine Learning (ICML)*. 2017.
  24. Cover TM, Thomas JA: *Elements of Information Theory*. edn 2. Wiley-Interscience; 2006.
  25. Schmidhuber J: **Driven by compression progress: a simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes.** *J SICE* 2009, **48**:21-32.
  26. Singh SP, Barto AG, Chentanez N: **Intrinsically motivated reinforcement learning.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2004:1281-1288.
  27. Houthoofd R, Chen X, Duan Y, Schulman J, De Turck F, Abbeel P: **VIME: variational information maximizing exploration.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2016:1109-1117.
  28. Mohamed S, Rezende D: **Variational information maximisation for intrinsically motivated reinforcement learning.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2015:1-9.
  29. Sorg J, Singh S, Lewis RL: **Variance-based rewards for approximate Bayesian reinforcement learning.** *Conference on Uncertainty in Artificial Intelligence (UAI)*. 2010:564-571.
  30. Burda Y, Edwards H, Storkey A, Klimov O: **Exploration by random network distillation.** *International Conference on Learning Representations (ICLR)*. 2019.
  31. Burda Y, Edwards H, Pathak D, Storkey A, Darrell T, Efros AA: **Large-scale study of curiosity-driven learning.** *International Conference on Learning Representations (ICLR)*. 2019.
  32. Haber N, Mrowca D, Wang S, Fei-Fei L, Yamins DL: **Learning to play with intrinsically-motivated self-aware agents.** *Conference on Neural Information Processing Systems (NeurIPS)*. 2018.
  33. Pathak D, Agrawal P, Efros AA, Darrell T: **Curiosity-driven exploration by self-supervised prediction.** *International Conference on Machine Learning (ICML)*. 2017.
  34. Gittins J, Glazebrook K, Weber R: *Multi-Armed Bandit Allocation Indices*. edn 2. John Wiley & Sons, Ltd.; 2011.
  35. Gittins JC: **Bandit processes and dynamic allocation indices.** *J R Stat Soc Ser B* 1979, **41**:148-177.
  36. Russo DJ, Van Roy B, Kazerouni A, Osband I, Wen Z: *A Tutorial on Thompson Sampling*. Foundations and Trends in Machine Learning; 2018:1-96.
  37. Thompson W: **On the likelihood that one unknown probability exceeds another in view of the evidence of two samples.** *Biometrika* 1933, **25**:285-294.
  38. Standing L: **Learning 10,000 pictures.** *Q J Exp Psychol* 1973, **25**:207-222.
  39. Brady TF, Konkle T, Alvarez GA, Oliva A: **Visual long-term memory has a massive storage capacity for object details.** *Proc Natl Acad Sci U S A* 2008, **105**:14325-14329.
  40. Bogacz R, Brown MW: **Comparison of computational models of familiarity discrimination in the perirhinal cortex.** *Hippocampus* 2003, **13**:494-524.
  41. Brown MW, Banks PJ: **In search of a recognition memory engram.** *Neurosci Biobehav Rev* 2015, **50**:12-28.
  42. Fahy FL, Riches IP, Brown MW: **Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex.** *Exp Brain Res* 1993, **96**:457-472.

43. Li L, Miller EK, Desimone R: **The representation of stimulus familiarity in anterior inferior temporal cortex.** *J Neurophysiol* 1993, **69**:1918-1929.
44. Xiang JZ, Brown MW: **Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe.** *Neuropharmacology* 1998, **37**:657-676.
45. Desimone R: **Neural mechanisms for visual memory and their role in attention.** *Proc Natl Acad Sci U S A* 1996, **93**:13494-13499.
46. Meyer T, Rust NC: **Single-exposure visual memory judgments are reflected in inferotemporal cortex.** *eLife* 2018, **7**:e32259.  
This paper demonstrated the plausibility of IT repetition suppression as a novelty signal by illustrating that a positively weighted linear read-out of IT responses could account for remembering and forgetting behavior as a function of time since an image was viewed.
47. DiCarlo JJ, Zoccolan D, Rust NC: **How does the brain solve visual object recognition?** *Neuron* 2012, **73**:415-434.
48. Grill-Spector K, Henson R, Martin A: **Repetition and the brain: neural models of stimulus-specific effects.** *Trends Cogn Sci* 2006, **10**:14-23.
49. McMahon DB, Olson CR: **Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming.** *J Neurophysiol* 2007, **97**:3532-3543.
50. Vogels R: **Sources of adaptation of inferior temporal cortical responses.** *Cortex* 2016, **80**:185-195.
51. Zhou J, Benson NC, Kay KN, Winawer J: **Compressive temporal summation in human visual cortex.** *J Neurosci* 2018, **38**:691-709.
52. Lim S, McKee JL, Woloszyn L, Amit Y, Freedman DJ, Sheinberg DL, Brunel N: **Inferring learning rules from distributions of firing rates in cortical neurons.** *Nat Neurosci* 2015, **18**:1804-1810.
53. Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T: **Neural repetition suppression reflects fulfilled perceptual expectations.** *Nat Neurosci* 2008, **11**:1004-1006.
54. Grotheer M, Kovacs G: **Repetition probability effects depend on prior experiences.** *J Neurosci* 2014, **34**:6640-6646.
55. Yildirim I, Siegel M, Tenenbaum JB: **Physical object representations for perception and cognition.** In *The Cognitive Neurosciences*, edn 6. Gazzaniga M, Poeppel MIT press; In press.
56. Hong H, Yamins DL, Majaj NJ, DiCarlo JJ: **Explicit information for category-orthogonal object properties increases along the ventral stream.** *Nat Neurosci* 2016, **19**:613-622.  
This paper demonstrated the robust and easily accessible representations of IT populations for properties beyond object identity, including object position.
57. Sawamura H, Orban GA, Vogels R: **Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm.** *Neuron* 2006, **49**:307-318.
58. De Baene W, Vogels R: **Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials.** *Cereb Cortex* 2010, **20**:2145-2165.
59. Schultz W: **Predictive reward signal of dopamine neurons.** *J Neurophysiol* 1998, **80**:1-27.
60. Kakade S, Dayan P: **Dopamine: generalization and bonuses.** *Neural Netw* 2002, **15**:549-559.
61. Yang SC-H, Wolpert DM, Lengyel M: **Theoretical perspectives on active sensing.** *Curr Opin Beh Sci* 2016, **11**:100-108.