Behavioral and Brain Sciences

cambridge.org/bbs

Open Peer Commentary

Cite this article: Zhou D, Bornstein AM. (2024) Expanding horizons in reinforcement learning for curious exploration and creative planning. *Behavioral and Brain Sciences* 1–3. doi:10.1017/S0140525X23003394

Commentaries Accepted: 4 December 2023

*Corresponding author.

CAMBRIDGE UNIVERSITY PRESS

Expanding horizons in reinforcement learning for curious exploration and creative planning

Dale Zhou[a,b*] and Aaron M. Bornstein[b,c]

[a]Neurobiology and Behavior, 519 Biological Sciences Quad, University of California, Irvine, CA, USA; [b]Center for the Neurobiology of Learning and Memory, Qureshey, Research Laboratory, University of California, Irvine, CA, USA and [c]Department of Cognitive Sciences, 2318 Social & Behavioral Sciences Gateway, University of California, Irvine, CA, USA.

dale.zhou@uci.edu
https://dalezhou.com
aaron.bornstein@uci.edu
https://aaron.bornstein.org/

Abstract

Curiosity and creativity are expressions of the trade-off between leveraging that with which we are familiar or seeking out novelty. Through the computational lens of reinforcement learning, we describe how formulating the value of information seeking and generation via their complementary effects on *planning horizons* formally captures a range of solutions to striking this balance.

Ivancovsky et al. propose fruitful connections between curiosity and creativity under an exploration–exploitation trade-off. The explore–exploit trade-off is the decision between a familiar option with known value and an unfamiliar option with unknown or uncertain value (Addicott, Pearson, Sweitzer, Barack, & Platt, 2017). Choosing unfamiliar options is risking time, energy, and foregone reward in return for information (Rubin, Shamir, & Tishby, 2012).

These ideas have history in reinforcement learning. For example, novelty-seeking is important to prevent failures of learning where subpar solutions are settled on prematurely (Fox, Pakman, & Tishby, 2015). Despite the benefits of novelty-seeking, seeking novel information can also carry a high cost when forgoing familiar opportunities and accruing a burdensome amount of information (Wilson, Bonawitz, Costa, & Ebitz, 2021). Thus, one must manage costs by taking "sensible risks" which balance exploring to learn novel information about the environment with accruing increasingly complex information for different tasks at hand (Sternberg & Lubart, 1996). One way to encourage taking on these risks for exploration is to use heuristics which locally track what has and has not been seen (Tang *et al.*, 2017; Wittmann, Bunzeck, Dolan, & Du¨zel, 2007; Wittmann, Daw, Seymour, & Dolan, 2008). By contrast, preferring familiarity can manifest as a form of perseverative information seeking that was associated with deprivation curiosity (Lydon-Staley, Zhou, Blevins, Zurn, & Bassett, 2021), a drive to reduce uncertainty and acquire missing information (Kashdan *et al.*, 2018; Litman, 2008). This preference for familiarity has been seen as prevalent in people with greater depressed mood and anxiety (Zhou *et al.*, 2023), and may be an important heuristic strategy to reduce uncertainty for better reliability of future-oriented decisions (Harhen & Bornstein, 2023; Jiang, Kulesza, Singh, & Lewis, 2015). However, in large environments, such local heuristics are impoverished, particularly when higher-order associations are needed for planning. This need for richer measurements motivates the use of network science tools to formalize both local and global relationships as internal representations of the environment (Yoo, Bornstein, & Chrastil, 2023; Zhou, Lydon-Staley, Zurn, & Bassett, 2020). Thus, we propose expansions of the novelty-seeking model using reinforcement learning approaches to exploration and network science perspectives on information complexity and compression.

Ivancovsky et al. rightly note that curiosity and creativity must involve a dynamic policy of behavior that adaptively alternates between modes of exploration and exploitation. Reinforcement learning approaches reveal what behavior pattern, or policy, is appropriate for a given task and environment, for instance adapted to the sparsity of rewarding solutions (Gershman & Niv, 2015). To this end, the reinforcement learning approach of Harada (2020) was described. However, notably this paper reported that divergent and convergent thinking measures of creativity and the personality trait of openness to experience (a proxy for being "inventive/curious") were *not* robustly associated to exploration and exploitation behavior based on model-free reinforcement learning (Harada, 2020). This finding highlights the need for understanding creativity via more sophisticated models of the value of exploration.

The value of information is sometimes treated as a simple heuristic for predisposing choices toward exploration (Gottlieb, Oudeyer, Lopes, & Baranes, 2013), but the value can also be formally expanded as the change in future expected value that results from increasing certainty

over representations of the environment and sequence of choices (Kaelbling, Littman, & Cassandra, 1998). These planning and policy iteration approaches aim for more global knowledge about the environment, and thereby differ from the local count-based reward functions to encourage exploration (Masís, Chapman, Rhee, Cox, & Saxe, 2023; Oudeyer & Kaplan, 2007; Tang *et al.*, 2017; Wittmann *et al.*, 2008). Here we focus on approaches that balance the increased long-run discounted expected value of knowledge with the cost of sampling (exploration) (Kaelbling *et al.*, 1998). To this end, the focus of choices shifts from an explore-or-exploit distinction to the iterative improvement of knowledge of the environment by testing predictions and simulations of future outcomes according to a given action policy (Gruber & Ranganath, 2019; Wilson, Wang, Sadeghiyeh, & Cohen, 2020).

We describe two areas of future research. First, creative insights can emerge from expanded planning horizons. Planning is commonly implemented as a search over a decision tree, wherein expanded horizons entail a deeper search in the tree. When the internal representation of information about the causal structure of the environment is accurate, longer planning horizons are useful. However, when the representation is incomplete, a smaller planning horizon compresses the policy space and prevents overfitting to past observations (Jiang *et al.*, 2015). Humans can search over more complex structures in knowledge representations (Yoo *et al.*, 2023). That knowledge may be more modular and compressible, allowing for the grouped representation of a more diverse chain of actions (Lai & Gershman, 2021; Momennejad, 2020; Patankar *et al.*, 2023; Schapiro, Rogers, Cordova, Turk-Browne, & Botvinick, 2013; Stachenfeld, Botvinick, & Gershman, 2017). The ability to use more complex knowledge structures may involve a spatial-like ability to navigate those structures (Rmus, Ritz, Hunter, Bornstein, & Shenhav, 2022), as well as a metacognitive ability to balance knowledge uncertainty with deeper planning. Indeed, a form of mental navigation that spans diverse spaces has been proposed to be linked with both creativity and curiosity (Aru, Drüke, Pikamäe, & Larkum, 2023; Eysenbach, Gupta, Ibarz, & Levine, 2018; Zhou *et al.*, 2023). Although such diversity and depth can decrease knowledge uncertainty, it comes at the cost of time and computational resources to accrue and update information. Computational cost motivates the next direction of research.

Second, creatively recombining knowledge benefits from unlearning or updating outdated knowledge. This form of creativity complements a type of curiosity that is characterized by deconstructing and rebuilding current structures (Zurn, 2021). When an agent seizes onto a supposedly optimal choice that is actually suboptimal, future resources must be used to unlearn those experiences (Fox *et al.*, 2015). This is precisely a problem that deprivation curiosity can exacerbate (Kruglanski & Webster, 2018; Zedelius, Gross, & Schooler, 2022). A solution to this problem involves aiming for simpler, compressed policies by chunking actions (Lai & Gershman, 2021). Compression involves smartly discarding some information to efficiently redescribe the information, such as by describing an elephant and a chicken with one joint description rather than describing each alone (Cover & Thomas, 1991; Mack, Preston, & Love, 2020). In order to modulate the planning horizon, policies could be compressed to increase certainty, albeit over an impoverished model. This idea is related to strategically decomposing, aggregating, and reducing sequences of actions into a hierarchy of "options" (Botvinick, Niv, & Barto, 2009; Sutton, Precup, & Singh, 1999) to balance the growing cost of planning (Botvinick, 2012; Correa, Ho, Callaway, Daw, & Griffiths, 2023). The idea also relates to a computational form of curiosity that involves improving prediction of expected long-term value (Gruber & Ranganath, 2019; Schmidhuber, 2008). Prediction is related to compression because the best compression is the true data generating model, and the true data generating model is the most predictive (Shannon, 1948). Notably, neural activity has been measured to be most compressed in the default-mode network (Mack *et al.*, 2020; Zhou *et al.*, 2022), a network of regions central to the proposed novelty-seeking model. Default-mode activity is also associated with the simulation of hypothetical episodes (Schacter & Addis, 2007) and the replay of episodic memories (Schapiro, McDevitt, Rogers, Mednick, & Norman, 2018), which can help to plan or update actions from new experiences (Kauvar, Doyle, Zhou, & Haber, 2023; Wilson *et al.*, 2020).

In conclusion, curiosity could be thought of computationally as actions taken to justify the expansion one's planning horizon. The consequent cost of increased complexity can be managed by creatively compressing action policies, which further support the pursuit of long-term goals.

Financial support. D. Z. acknowledges funding from the George E. Hewitt Foundation for Medical Research. A. M. B. acknowledges funding from NINDS R01NS119468 (PI: E.R. Chrastil) and NIMH R01MH128306 (PI: M.A. Yassa).

Competing interests. None.

References

Addicott, M. A., Pearson, J. M., Sweitzer, M. M., Barack, D. L., & Platt, M. L. (2017). A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology*, *42*(10), 1931–1939.

Aru, J., Drüke, M., Pikamäe, J., & Larkum, M. E. (2023). Mental navigation and the neural mechanisms of insight. *Trends in Neurosciences*, *46*(2), 100–109.

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology*, *22*(6), 956–962.

Botvinick, M. M., Niv, Y., & Barto, A. G. (2009). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, *113*(3), 262–280.

Correa, C. G., Ho, M. K., Callaway, F., Daw, N. D., & Griffiths, T. L. (2023). Humans decompose tasks by trading off utility and computational cost. *PLoS Computational Biology*, *19*(6), e1011087.

Cover, T. M., & Thomas, J. A. (1991). Rate distortion theory. *Elements of Information Theory*, 336–373. Q3

Eysenbach, B., Gupta, A., Ibarz, J., & Levine, S. (2018). Diversity is all you need: Learning skills without a reward function. *arXiv preprint arXiv*:1802.06070.

Fox, R., Pakman, A., & Tishby, N. (2015). Taming the noise in reinforcement learning via soft updates. *arXiv preprint arXiv*:1512.08562.

Gershman, S. J., & Niv, Y. (2015). Novelty and inductive generalization in human reinforcement learning. *Topics in Cognitive Science*, *7*(3), 391–415.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*(11), 585–593.

Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (pace) framework. *Trends in Cognitive Sciences*, *23*(12), 1014–1025.

Harada, T. (2020). The effects of risk-taking, exploitation, and exploration on creativity. *PLoS ONE*, *15*(7), e0235698.

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences*, *120*(13), e2216524120.

Jiang, N., Kulesza, A., Singh, S., & Lewis, R. (2015). The dependence of effective planning horizon on model accuracy. In *Proceedings of the 2015 international conference on autonomous agents and multiagent systems* (pp. 1181–1189).

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134.

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). The five-dimensional curiosity scale: Capturing the bandwidth

of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality*, *73*, 130–149.

Kauvar, I., Doyle, C., Zhou, L., & Haber, N. (2023). Curious replay for model-based adaptation.

Kruglanski, A. W., & Webster, D. M. (2018) Motivated closing of the mind: "seizing" and "freezing". *The Motivated Mind*, 60–103.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In *Psychology of learning and motivation* (Vol. 74, pp. 195–232). Elsevier.

Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and Individual Differences*, *44*(7), 1585–1595.

Lydon-Staley, D. M., Zhou, D., Blevins, A. S., Zurn, P., & Bassett, D. S. (2021). Hunters, busybodies and the knowledge network building associated with deprivation curiosity. *Nature Human Behaviour*, *5*(3), 327–336.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications*, *11*(1), 46.

Masís, J., Chapman, T., Rhee, J. Y., Cox, D. D., & Saxe, A. M. (2023). Strategically managing learning during perceptual decision making. *Elife*, *12*, e64978.

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics*, *1*, 6.

Patankar, S. P., Zhou, D., Lynn, C. W., Kim, J. Z., Ouellet, M., Ju, H., … Bassett, D. S. (2023). Curiosity as filling, compressing, and reconfiguring knowledge networks. *Collective Intelligence*, *2*(4), 26339137231207633.

Rmus, M., Ritz, H., Hunter, L. E., Bornstein, A. M., & Shenhav, A. (2022). Humans can navigate complex graph structures acquired during latent learning. *Cognition*, *225*, 105103.

Rubin, J., Shamir, O., & Tishby, N. (2012). Trading value and information in mdps. *Decision making with imperfect decision makers* (pp. 57–74).

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 773–786.

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, *9*(1), 3920.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492.

Schmidhuber, J. (2008). Driven by compression progress: A simple principle explains essential aspects of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Workshop on anticipatory behavior in adaptive learning systems* (pp. 48–76).

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653.

Sternberg, R. J., & Lubart, T. I. (1996). Investing in creativity. *American Psychologist*, *51* (7), 677.

Sutton, R. S., Precup, D., & Singh, S. (1999). Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, *112*(1–2), 181–211.

Tang, H., Houthooft, R., Foote, D., Stooke, A., Xi Chen, O., Duan, Y., … Abbeel, P. (2017). # exploration: A study of count-based exploration for deep reinforcement learning. *Advances in Neural Information Processing Systems*, *30*.

Wilson, R., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences*, *38*, 49–56.

Wilson, R., Wang, S., Sadeghiyeh, H., & Cohen, J. D. (2020). Deep exploration as a unifying account of explore-exploit behavior.

Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Du¨zel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuroimage*, *38*(1), 194–202.

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008). Striatal activity underlies novelty-based choice in humans. *Neuron*, *58*(6), 967–973.

Yoo, J., Bornstein, A., & Chrastil, E. R. (2023). Cognitive graphs: Representational substrates for planning.

Zedelius, C. M., Gross, M. E., & Schooler, J. W. (2022). Inquisitive but not discerning: Deprivation curiosity is associated with excessive openness to inaccurate information. *Journal of Research in Personality*, *98*, 104227.

Zhou, D., Kim, J. Z., Pines, A. R., Sydnor, V. J., Roalf, D. R., Detre, J. A., … Bassett, D. S. (2022). Compression supports low-dimensional representations of behavior across neural circuits. *bioRxiv*, 2022–11.

Zhou, D., Lydon-Staley, D. M., Zurn, P., & Bassett, D. S. (2020). The growth and form of knowledge networks by kinesthetic curiosity. *Current Opinion in Behavioral Sciences*, *35*, 125–134.

Zhou, D., Patankar, S., Lydon-Staley, D. M., Zurn, P., Gerlach, M., & Bassett, D. S. (2023). Architectural styles of curiosity in global Wikipedia mobile app readership. *PsyArXiv*.

Zurn, P. (2021). Curiosity: An affect of resistance. *Theory & Event*, *24*(2), 611–617.