## 18 Jan 2023, How might curiosity-driven learning AI/ML warfighting systems help to probe and exploit opportunities in future warfare?


There is progress in the development of AI/ML systems that self-adapt to evolving contingencies using a learning method referred to as 'reinforcement learning' (RL). The most common learning approach employed by AI/ML-RL systems is to learn based on external or extrinsic rewards. Such systems operate with goals for maximizing long-term reward. This RL approach has proven effective for AI/ML-RL systems to play complex games (in some cases even better than human experts) and for robotic systems learning how to move and perform tasks. In each example, the AI/ML-RL systems are focused on pre-determined problems offering extrinsic rewards for expected optimal actions that have already been previously found to be effective. The extrinsic rewards are defined by designers and developers of AI/ML-RL systems to serve as reinforcers for action when triggered by sensors detecting their presence in the environment.

Overall, extrinsic reward forms of AI/ML-RL systems are effective for adaptations to well-defined problems using rewards that are effective on the basis of pre-trained and expected acquired knowledge. Extrinsic forms of AI/ML-RL systems work well for recurring problems with known cause-effect chains (e.g., consequences of action taken can be reliably anticipated). But, they perform poorly for complex problems that are unpredictable and non-causal. Under such circumstances, actions anticipated and taken by AI/ML-RL systems can produce unintended consequences and failures as the complex nature unfolds in time. Whereas, humans can be motivated from curiosity as an 'intrinsic reward' function probing why predictively effective actions might have failed. Probing can lead to better insights for improving adaptations and recognizing unplanned opportunities in flow with the unfolding nature of what is being encountered.

There is no similar means offered by an extrinsic reward form of an AI/ML-RL system to employ human-like artificial curiosity to generate intrinsic rewards for learning when predicted optimal actions fail. What is needed is AI/ML-RL system behavior driven by intrinsic objectives and rewards, rather than by hard-wired programs, external supervision, or external rewards (see Lecun, 2022). For an AI/ML-RL system to intrinsically reward itself it needs to probe, discover, and exploit strategies to complex problems not previously conceived or anticipated by its designers and developers. There are two forms of intrinsic reward functions that have been proposed for AI/ML-RL systems: 1) those that focus on rewarding novelty, and 2) those that focus on finding gaps in the system's model of the world to reduce prediction error (see Dubey & Griffiths, 2020).

Both forms of intrinsic reward functions offer unique prospects for usage of curiosity-driven learning (CDL) AI/ML-RL warfighting systems. The ultimate self-adapting goal of CDL AI/ML-RL warfighting systems would be to learn optimal actions to complete target tasks by maximizing the expected rewards from the actions taken in the environment. To better highlight the importance and mechanisms supporting CDL AI/ML-RL warfighting systems it is worthwhile to briefly dive into the nature of curiosity and CDL related to humans. Much of neural development can be understood as learning contingencies about the world and how to effectively act on the world (Friston, et.al., 2017). Such learning rests primarily on intrinsically motivated curious behavior. Humans attempt to resolve uncertainty using an adaptive mental world model with attempts to fine tune the model to maximize optimal actions in the future. For humans this makes the world both interesting and exploitable using internal rewards as motivators for curious behavior (Still & Precup, 2012). Thus, even in the absence of external rewards humans can be internally motivated for employing curious behavior to explore and probe their world.

Research on human curiosity suggests an inverted U-shaped function influenced by levels of knowledge and problem difficulty. Interestingly, people showing highest curiosity tend to possess moderate knowledge about encountered problems that are moderately difficult. This interesting aspect of human curiosity, involving the interplay of levels of knowledge and problem difficulty, offers insights into the dynamic nature and use of curious behavior for managing knowledge gaps. If the problem or challenge offers predictable patterns over time then CDL behavior is likely to be effective. Of course, this approach can be highly problematic if past and future occurrences of a problem are independent of each other! Thus, if the problem presents unpredictable patterns then CDL exploration and probing for optimal actions might make better sense (see Dubey & Griffiths, 2020). And, if the problem presents as chaotic then radical re-purposing of an existing mental model of the world may be where CDL probing ultimately takes a person.

All of this said, human CDL behavior observed in the face of unpredictable white-waters of emergent and unfolding complex events prioritizes probing for ways to reduce uncertainty. Fortunately for humans, in the face of complex, chaotic circumstances, initial use of CDL prediction-error reduction, can shift to another approach. Under highly uncertain circumstances it might be more appropriate to employ a CDL probing approach to better make sense of a complex, unpredictable problem. In fascinating ways, people can attain insight from probing using just a handful of observations, which are solicited through curious behavior (Friston, et.al., 2017). Wisdom to discern when to switch and employ appropriate CDL approaches called for by the nature of a problem matters for effective human learning. Might the same discernment capability be reasonable to pursue in AI/ML-RL warfighting systems for optimal actions in future warfare? How might this be possible?

Deep learning offers one possible solution. Progress with deep learning, offering means to address a variety of problems using weights involving large artificial neural networks that learn through experience, has made it possible to scale up external rewards-based AI/ML-RL systems for more complicated problems that were previously considered intractable (Dubey & Griffiths, 2020, p. 119). For example, Lempert (2019), highlights AI progress with engaging in problems of deep uncertainty using Robust Decision Making (RDM) employing problem-specific algorithmic models. Admittedly, progress with external reward-based AI/ML-RL systems is helping to increase the variety of anticipated problems a warfighting system might be capable of handling. But, as highlighted above, dependence on external rewards for learning, adapting, and determining optimal actions might prove insufficient in the face of unpredictable problems and complex events. Is there an alternative to the use of external rewards? Fortunately, progress is being made to offer AI/ML-RL systems with CDL functional means to engage in appropriate artificial curiosity-driven behavior in the face of highly uncertain and unpredictable events evoked on the basis of internal rewards. Much of the effort centers around the development of processes that can account for surprise in support of uncertainty-resolving behaviors. Specifically, so-called 'CDL autonomous learning agents' are being designed and evaluated that can spontaneously learn rules to govern actions making use of limited and ambiguous sensory evidence characteristic with surprising events (Friston, et.al., 2017).

What is interesting with such efforts involves deep learning architectures making use of generative models trained with self-supervised learning with which the system can predict the consequences of its actions and plan ahead for optimal actions. The models are optimized for inferencing to find simpler, more critical aspects of a situation as it unfolds with a variety of contingencies (see Sun, et.al., 2022, p. 18). Simpler explanations for contingencies are inherently more generalizable in service to efficient inferencing of optimal actions for novel, unpredictable, and highly uncertain problem contexts. This capability can help guide an AI/ML-RL system with forms of probing akin to curiosity in humans necessary for learning new behaviors and adapting to dynamic situations (Ibid, 2022; Oudeyer, 2018). The more uncertain and dynamic the situation, the more proximate a behavioral choice must be. This not only holds true for human adaptability to the complexity of reality but also is desired for adaptive machine behavior in service to humans.

Promising effort with artificial CDL is focused on developing curiosity mechanisms aimed to improve subjective estimation of complex world states (used for adapting and proximate determination of optimal actions under changing circumstances; referred to as active inferencing). World states are represented in internal deep learning neural network model structures, representing dynamic world states (think: operating environments), used for inferencing, self-diagnosis, and adaptive action by a CDL AI/ML-RL system. Refinement of subjective estimations made possible with curiosity mechanisms help with assessing gaps about the systems' internal model structure necessary to improve inferencing and supporting learning about its' own structural efficacy for informing self-directed improvements in dynamic situations (Ibid, 2018). These advances are making it possible for a CDL AI/ML-RL system to learn, probe, and operate using context-sensitive contingencies, associated with complex reality, by self-directing adaptations and resetting proximate objectives close enough in hand to unfolding events to be feasible. Such systems make use of active inferencing, that assimilates previous experience for attentional action or 'curiosity activation' for reducing uncertainty and improving by managing its own knowledge gaps (see Sun, et.al., 2022).

Basically, this capability enables CDL AI//ML-RL systems to better determine optimal actions for complex and dynamic situations. There is still much work to be done. Future systems need to make determinations with means to also take into account that any action pursued is very likely to impact the evolvement of hidden states of a complex event. This calls for continued probing and sense-making of changing circumstances to revisit self-supervised adjustments for improved inferencing. The math behind these advancements involves the use of a generative model to specify joint probability of outcomes or consequences and their latent or hidden causes. This line of effort follows the research on what is known about CDL and how people use their knowledge of real-world probabilistic assessments. Such assessments inform search behavior of viable optimal actions while also enabling extra means for CDL AI/ML-RL warfighting systems to test plausible hypotheses about internal generative models. Ultimately, such systems will discover gaps in existing knowledge structures calling for the identification of rules necessary to address the gaps by self-directed adaptations. Progress is also being made with CDL AI/ML-RL systems to address knowledge gaps using just a handful of observations similar to humans (bid, 2017). Where might this progress lead for future CDL AI/ML-RL warfighting systems?

As highlighted above, it is important for future CDL AI/ML-RL warfighting systems to employ attentional  action to activate curiosity-driven behaviors necessary for self-supervised assessment and management of knowledge gaps in its' own internal model structure. This capability enables inferencing about possible optimal actions in the face of complexity. While this capability is important, future CDL AI/ML-RL warfighting systems need to be able to perform without excessive computational costs on a short time scale necessary for survival and exploitation of opportunities in warfare. Also, future CDL AI/ML-RL warfighting systems must not become vulnerable or easily distracted by task-irreverent novelty and unpredictable noise in service to maintaining an optimal balance between probing and exploitation.

In addition to prospects offered with the use of CDL AI/ML-RL systems in warfighting there are also benefits expected with their use as recommender systems. For example, a CDL AI/ML-RL system might be employed in support of JADC2 to help probe and identify knowledge gaps and emerging patterns across several domains to address unforeseen exploitable opportunities in warfare. Imagine future CDL AI/ML-RL warfighting and recommender systems offering means to self-generate a goal not originally anticipated to address novelty, adapt to unpredictability, and exploit unforeseen opportunities. There is every expectation future CDL AI/ML-RL warfighting systems can offer improved means to recover from failed actions while also taking initiative to probe for alternative optimal actions for successful mission outcomes aligned under commander's intent.

### From the AI S&C Bibliography


Dubey, R., & Griffiths, T. L. (2020). Understanding exploration in humans and machines by formalizing the function of curiosity. *Current Opinion in Behavioral Sciences*, 35, 118–124.

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural computation*, 29(10), 2633–2683.

Gottlieb, J., & Oudeyer, P.-Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 19(12), 758–770.

Lecun, Y. (2022). A path towards autonomous machine intelligence. *Open Review*, 62.

Lempert, R.J. (2019). Robust decision making. In V. Marchau, W. Walker, P. Bloeman, & S. Popper (eds.), *Decision Making Under Uncertainty: From Theory to Practice*, Springer.

Oudeyer, P.-Y. (2018). Computational theories of curiosity-driven learning. arXiv preprint arXiv:1802.10546.

Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Bioscience*s, 131, 139-148.

Sun, C., Qian, H., & Miao, C. (2022). From psychological curiosity to artificial curiosity: Curiosity-driven learning in *Artificial Intelligence Tasks*. arXiv preprint arXiv:2201.08300.