Learning Fast and Slow: A Redux of Levels of Learning in General Autonomous Intelligent Agents

Shiwali Mohan[1] **and John E. Laird**[2]

[1] SRI International, Palo Alto, California, USA
[2] Center for Integrated Cognition, Ann Arbor, Michigan, USA
shiwali.mohan@sri.com, laird@cic.iqmri.org

The world that general autonomous intelligent agents (e.g., humans) operate in is complex, dynamic, partially observable, and unknown. We must continually react to their environment, focusing our computational resources on making the best decision for the current situation using all our available knowledge. We need to learn everything we can from our experience, building up our knowledge so that we are prepared to make the best decisions in the future. We have evolved to be effective and efficient learners to maximize our available knowledge.

We revisit the thesis we proposed originally in Laird and Mohan (2018). Analogous to dual-process theory (Wason and Evans 1974), popularized by Kahneman (2011) as System 1 and System 2, our thesis posits that learning can be split into two levels in human-like, generally intelligent agents. These levels are inspired by research across neuroscience, psychology, and cognitive architecture (Laird, Lebiere, and Rosenbloom 2017). Level 1 (or L1) learning encompasses *architectural learning mechanisms* that are innate, automatic, and effortless, such as the temporal difference update mechanism in reinforcement learning (Van Hasselt, Guez, and Silver 2016) or optimization of convolutional neural network (Gu et al. 2018) architectures. These processes are similar to System 1 processes that are autonomous, implicit, and always on. Level 2, or L2, encompasses *deliberate learning strategies* that are realized through knowledge and controlled by an agent. These strategies can be thought of as learning tasks that the agent deliberately adopts and pursues using System 2 reasoning. Like System 2 reasoning tasks, L2 learning competes with performance tasks for mental (and physical) resources and depends on L1 mechanisms for knowledge acquisition. A simple example in humans is deciding to rehearse a phone number to memorize it explicitly. Deliberately repeating the number several times aloud (or to oneself) creates experiences that are consolidated by automatic L1 memory mechanisms, making the number available for later recall. L2 strategies can include additional learning mechanisms and use the L1 mechanisms to extract regularities and record knowledge structures from the generated experiences.

Our thesis is based on our experiences in building artificial agents that have complex learning behaviors. Interactive task learning (ITL: Mohan and Laird 2014; Laird et al. 2017; Kirk and Laird 2016; Mininger and Laird 2018; Kirk and Laird 2019) enables an agent to learn from natural human teaching. Open-world learning (OWL: Mohan et al. 2023; Piotrowski et al. 2023) enables an agent to learn in a continually evolving, unknown world. ITL and OWL challenges necessitate online learning interleaved with task performance. Consequently, the agent must autonomously identify when its current knowledge is insufficient to handle the task and commence efforts for learning. These problems pose a uniquely different challenge than the classical machine learning (ML) setup of training and testing, where the ML designer evaluates a learning system's performance and trains it. Further, ITL and OWL motivate incremental learning that exploits and builds upon what was learned previously without full retraining, common in ML literature.

For ITL (ROSIE) and OWL (HYDRA), we studied and developed L2 strategies that generate experiences for L1 mechanisms and enable incremental and continual learning in the agent. L2 strategies are reasoned about in the phases below.

1. In both ROSIE and HYDRA, the decision of applying an L2 strategy begins with with *identifying an opportunity to learn*. ROSIE relies on Soar's (Laird 2012) automatic ability to detect when knowledge is missing or conflicting, which generates a subgoal. In this subgoal, ROSIE can pursue additional reasoning or information gathering to make progress. HYDRA consists of novelty indicators that maintain explicit expectations over perceptions, transition dynamics, and performance quality when operating in an environment. A significant divergence of those from observations indicates an existence of novelty, causing the agent to devote more resources to learning. Generally, other opportunities may be identified because an external knowledge source (e.g., a more experienced agent) evaluates the learning agent's performance as deficient or the agent expects to be engaged in a future task where expert-level performance is important (e.g., participating in a rescue mission as a firefighter).
2. Next, is *selecting a learning strategy* in the face of other competing needs such as continuing to pursue the task. ROSIE pursues a straightforward strategy of acquiring task definitions through interactions with the instructor. It doesn't, yet, decide what the best way to learn is. In comparison, HYDRA characterizes which of its

decision-making components contributes to expectation-observation mismatch and selects a relevant learning strategy. For instance, if there is a significant disparity in its internal transition models and how the state evolves in the world, it chooses to repair them.

3. Upon selection, the agent *executes the learning strategy* to correct, adapt, or extend its knowledge. ROSIE leverages retrospective analysis, retrieving the instructed task performance from episodic memory and causally analyzing why that specific sequence of actions led to success. Through this causal analysis, it infers how it should sequence the actions to perform a specific task. HYDRA revises its transition dynamics models using a heuristics-based search process. The process is engaged when its internal simulation (available as a plan) is inconsistent with what it observes when it executes the plan in the world. The search process incrementally changes various adaptable parameters of its transition models until it finds a parameter value assignment (termed a repair) that makes its internal simulation consistent with what it observed.
4. Finally, the agent *uses the learned knowledge* to progress on its tasks. It may decide to test, verify, and monitor acquired knowledge to ensure its correctness and relevance. In both ROSIE and in HYDRA, acquired knowledge is immediately available for use by the agent for any future problem. As they apply learned knowledge, they may identify missing knowledge, inconsistencies, or novelties and consequently, start phase 1 over again.

Our thesis is relevant for intelligent, complex agents that have an extended existence and are expected to pursue a growing variety of tasks. An interesting class of agents is the recent iteration of agents developed using generative models such as large language models (LLMs). Such agents are designed within the training-testing paradigm and learn offline on curated benchmark datasets until a performance criterion is achieved. Adapting their behavior to novel tasks requires either fine-tuning or example-based prompting. While fine-tuning requires the system to be taken offline, example-based prompting can change the system's behavior online. However, the design of the prompt and subsequent refinements are controlled by an entity (a human designer) external to the learning system. It can be argued that the human designer plays the role of an L2 mechanism - they determine what prompt should be used to generate task-relevant behavior, evaluate system performance, and refine prompts when behavior diverges from the designer's expectations.

One can imagine future architectures built upon generative models where task-relevant learning is online, incremental, and autonomous (Laird 2012). These architectures would likely implement our L1/L2 thesis. The continual update of the underlying generative model (such as a transformer (Vaswani et al. 2017)) would correspond to L1 for such agents. L2 is a bit dicier for pure generative model systems and would require discovering ways of inducing the generative model to produce strategic learning behavior, possibly similar to how they can be induced to perform step-by-step reasoning with chain-of-thought and its siblings. Of course, by integrating the generative model as a component of a broader cognitive architecture or reasoning frameworks, other reasoning methods such as search, planning, or analogy could be responsible for generating the L2 strategies.

References

Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. 2018. Recent advances in convolutional neural networks. *Pattern recognition*, 77: 354–377.

Kahneman, D. 2011. *Thinking, Fast and Slow*. Macmillan.

Kirk, J. R.; and Laird, J. E. 2016. Learning General and Efficient Representations of Novel Games Through Interactive Instruction. *Advances in Cognitive Systems*, 4.

Kirk, J. R.; and Laird, J. E. 2019. Learning Hierarchical Symbolic Representations to Support Interactive Task Learning and Knowledge Transfer. In *In Proceedings of the International Joint Conference on Artificial Intelligence*, 6095–6102.

Laird, J.; Lebiere, C.; and Rosenbloom, P. 2017. A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. *AI Magazine*, 38(4).

Laird, J.; and Mohan, S. 2018. Learning Fast and Slow: Levels of Learning in General Autonomous Intelligent Agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Laird, J. E. 2012. *The Soar Cognitive Architecture*. MIT press.

Laird, J. E.; Gluck, K.; Anderson, J.; Forbus, K. D.; Jenkins, O. C.; Lebiere, C.; Salvucci, D.; Scheutz, M.; Thomaz, A.; Trafton, G.; et al. 2017. Interactive Task Learning. *IEEE Intelligent Systems*, 32(4): 6–21.

Mininger, A.; and Laird, J. E. 2018. Interactively Learning a Blend of Goal-Based and Procedural Tasks. In *In Proceedings of the Annual AAAI Conference on Artificial Intelligence*.

Mohan, S.; and Laird, J. E. 2014. Learning Goal-Oriented Hierarchical Tasks from Situated Interactive Instruction. In *In Proceedings of the Annual AAAI Conference on Artificial Intelligence*, 387–394.

Mohan, S.; Piotrowski, W.; Stern, R.; Grover, S.; Kim, S.; Le, J.; and De Kleer, J. 2023. A Domain-Independent Agent Architecture for Adaptive Operation in Evolving Open Worlds. *arXiv preprint arXiv:2306.06272*.

Piotrowski, W.; Stern, R.; Sher, Y.; Le, J.; Klenk, M.; deKleer, J.; and Mohan, S. 2023. Learning to Operate in Open Worlds by Adapting Planning Models. *International Conference on Autonomous Agents and Multiagent Systems*.

Van Hasselt, H.; Guez, A.; and Silver, D. 2016. Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 1.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wason, P. C.; and Evans, J. S. B. 1974. Dual processes in reasoning? *Cognition*, 3(2): 141–154.