AFRL-AFOSR-VA-TR-2021-0076

Levels of Learning in Natural and Artificial Agents

Laird, John
REGENTS OF THE UNIVERSITY OF MICHIGAN
503 THOMPSON ST
ANN ARBOR, MI, 48109
USA

07/15/2021
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
15-07-2021

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
15 Apr 2018 - 14 Apr 2021

4. TITLE AND SUBTITLE
Levels of Learning in Natural and Artificial Agents

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-18-1-0180

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
John Laird

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
REGENTS OF THE UNIVERSITY OF MICHIGAN
503 THOMPSON ST
ANN ARBOR, MI 48109
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2021-0076

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Throughout the history of psychology, there has been a recognition that there are multiple time-scales of decision making in human cognition. This includes early dual-process theories, but was recently popularized by Daniel Kahneman in his book, “Thinking Fast and Slow” (2011), where he proposed System 1 and System 2. We propose that it is useful to make a related distinction in learning for humans and intelligent autonomous agents based on how learning fits into the overall cognitive architecture. We initially define Level 1 (L1) and Level 2 (L2), where L1 are fixed, innate, automatic (architectural) learning mechanisms, and L2 are (knowledge-based) learning strategies that are controlled by the agent to create experiences such that L1 mechanisms can learn useful knowledge. The purpose of this project was to explore and refine these distinctions across both artificial and natural intelligent autonomous systems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			HAL GREENWALD
U	U	U	UU	8	**19b. TELEPHONE NUMBER** *(Include area code)* 703-588-8441

Standard Form 298 (Rev.8/98)
Prescribed by ANSI Std. Z39.18

AFOSR Final Report
Levels of Learning in Natural and Artificial Agents

Grant FA9550-18-1-0180

John E. Laird and Bryan Stearns, University of Michigan
Shiwali Mohan, PARC

July 13, 2021

Overview of Project

Throughout the history of psychology, there has been a recognition that there are multiple time-scales of decision making in human cognition. This includes early dual-process theories, but was recently popularized by Daniel Kahneman in his book, "Thinking Fast and Slow" (2011), where he proposed System 1 and System 2. We propose that it is useful to make a related distinction in learning for humans and intelligent autonomous agents based on how learning fits into the overall cognitive architecture. We initially define Level 1 (L1) and Level 2 (L2), where L1 are fixed, innate, automatic (architectural) *learning mechanisms*, and L2 are (knowledge-based) *learning strategies* that are controlled by the agent to create experiences such that L1 mechanisms can learn useful knowledge. The purpose of this project was to explore and refine these distinctions across both artificial and natural intelligent autonomous systems.

In pursuing this project, we have ended up with two major types of activities.

1. **Analyses of learning, performance, and architecture.**
    a. This document contains our initial analysis of the interaction between learning and performance, including the development of set of dimensions that distinguish different types of integration, and then a taxonomy of different approaches to integration using those dimensions. Our major focus has been on AI systems and how learning integrates with performance within an agent, including how the agent is embedded in its environment. We have generated multiple taxonomies and classification schemes that give a finer grain analysis than our initial Levels of Learning. These in turn have led to insights into why few AI systems employ Level 2 learning strategies. The most prevalent reasons are that AI learning systems are rarely embedded in agents with ongoing, long-term existence where the issues arise related to how learning interacts with and is integrated with performance systems. Moreover, they rarely have the metacognitive capabilities required for L2 strategies, which may also be a critical aspect missing in animals.

   b. That analysis led us to dig deeper into how Level 1 learning is integrated into a cognitive architecture, and how it interacts with the different types of data structures, metadata, and processes. The result was the development of an in-depth analysis and comparison of ACT-R and Soar, two Common Model cognitive architectures. This led to many insights into the interaction of data and metadata (and meta-process data) in these architectures, and a formalization of the major aspects of their processing. A draft of this paper is included as an addendum to this final report.

2. **AI agent development:**
   a. One of our early findings was that there was a lack of Level 2 learning in AI agents. In order to better understand the issues involved, we developed two AI agents that use both Level 1 and Level 2 learning. The first is one that attempts to model human task learning and transfer across a variety of tasks by reimplementing Niels Taatgen's PRIMS system that was implemented a variant of ACT-R, in Soar. Thus, this project involved not only studying how deliberate strategies (Level 2) can be used for task learning both in humans and artificial agents via Level 1 mechanisms, but it also provided an opportunity for an explicit comparison of ACT-R and Soar. This work constituted the Ph.D. thesis of Bryan Stearns, who completed his thesis on June 30, 2021, and a copy of that thesis is included with this report.
   b. The work on the second agent was to look at novel combinations/implementations of learning mechanism in a cognitive architecture and examine its impact on learning, both L1 and L2. Specifically, we extended the Soar cognitive architecture with a new learning mechanism based on the analogical reasoning and generalization in the Companions cognitive architecture - the Structure Mapping Engine (SME; Forbus et al. 2017) and the Sequential Analogical Generalization Engine (SAGE; McLure et al. 2015). This L1 mechanism supports learning of concept knowledge, as well as generalization of that knowledge. The resulting architecture was used to create AILEEN (Mohan et al. 2020), a cognitive system that learns new concepts through interactive experiences with a trainer in a simulated world. Our recent analysis (Ramaraj et al. 2021) of natural human teaching shows that human teachers rely on meta-cognitive feedback from a learning partner to adapt their teaching strategies. Our findings suggest that L2 strategies that leverage meta-cognitive analyses of own performance are critical for generally intelligent behavior in cognitive systems.

Below is a summary of our analyses of learning, performance, and architecture. The first section is a brief summary of our literature review of learning levels from the perspective of different disciplines: neuroscience, animal behavior, psychology, and artificial intelligence. Possibly the most important aspects of Level 1 and Level 2 learning is how they integrate with agent performance. Understanding the question led to the second and third sections. The second section attempts to identify the key dimensions of integration between learning and performance, with the third section then developing a taxonomy of different approaches to integration as those dimensions vary.

1. **Analysis of Learning Across Disciplines**

These three objectives blend in our analysis. Our emphasis, in terms of effort, has been focused mostly on artificial intelligence and cognitive psychology, because of a lack of findings in the other areas.

a. Animal behavior. We have found little to no evidence of Level 2 learning in non-human animals. For example, there is no evidence of deliberate self-driven practice. There is curiosity and play, which we interpret as an evolution-determined activity along the lines of Level 2, but not one where the animal does it with the *purpose* of improving its own future performance. There is also some evidence of experienced animals inducing younger animals to practice, but again, none where an animal acts deliberately to improve its own knowledge or skills. Deliberation, or explicit goal setting, has become an important criterion for Level 2 learning.

b. Neuroscience. Our investigations of neuroscience suggest that Level 2 is currently beyond the scope of study of the field. The learning mechanisms studied in neuroscience appear to be Level 1. One important observation is that not all Level 1 mechanisms lead to immediate learning, but some involve changes that take place over time.

c. Cognitive and educational psychology. Based on our literature review, we found recognition of a distinction between human deliberation for the sake of task performance and deliberation for the sake of learning. The metacognition literature in particular investigates an individual's awareness of their task performance or awareness of their own learning. The educational psychology literature is largely concerned with teacher-induced learning rather than deliberate, self-driven learning, but there is substantial work exploring the motivations for learning within students, contrasting the motivation to complete assignments (task performance) with the motivation to improve knowledge or expertise (deliberate, self-directed learning). In the developmental psychology literature, as would be expected, we found evidence of curiosity- and play-driven learning in infants and young children, and also of the gradual acquisition of self-directed, deliberate learning behaviors with age, but little or no explanation for the development of that ability. Overall, while there are many studies describing or modeling the separate phenomena of motivation, deliberation, and learning in humans, we have found little record in the psychology literature of attempts to explain the Level 2 learning phenomenon.

d. Artificial intelligence. We have done an extensive review of learning in AI systems including domains of interactive machine learning, active learning, reinforcement learning and deep learning as well as drawing from a review of over 84 cognitive architectures. Learning in AI is often categorized as based on the information available during learning, such as whether it is supervised, unsupervised, or reinforcement learning. Further distinctions are based on the underlying techniques/technologies used for processing and representing the learned knowledge, such as case-based reasoning, decision trees, explanation-based learning, neural networks, or genetic algorithms, among others. While another dimension is how the learned knowledge is used, such as for categorization, controlling problem solving (a policy), modeling the agent's actions and the environment, and so on. Our analysis is agnostic to these dimensions and instead focuses on how learning integrates with performance within an agent, including how the agent is embedded in its environment. Our goal was to develop a set of dimensions for this analysis that leads to a categorization of learning mechanisms, as described below.

2. Interactions Between Learning Performance: Dimensions of Variation

Our analysis led us to identifying the following dimensions of variation in learning processes and their embedding with performance in an agent:

1. Timing. [When] Is the frequency and duration of learning externally or internally defined? That is, is the timing of learning controlled by a third party, or is it determined on-line within the agent as it operates within its environment?
2. Direction. [Explicit goal deliberation] Is the process's learning goal externally or internally defined? That is, is the goal forced as an implicit property of its design, or does the process direct its own explicit learning goals?
3. Data collection. [What Information] Is the information used for learning externally or internally defined? That is, does a third party select or curate the content for learning, or is the content selected by the process on-line?
4. Learning execution. [Invoked External Learning Processes] Is the learning process itself externally or internally executed with respect to the process? That is, does the process force other processes to externally perform its learning

goals, or does it operate as its own process, though potentially allowing or expecting other processes to operate independently and automatically. Note: What is true when a human externally forces a learning process is also true when a learning process externally forces another learning process. Designing a process to perform learning with externally controlled timing interferes with that process's ability to function automatically.

3. Interactions Between Learning and Performance: Taxonomy of Approaches

From the dimensions described above, we develop a set of learning approaches that are defined by common combinations of different values in specific dimensions. This is essentially an attempt to create a taxonomy of learning approaches:

1. Batch learning followed by performance. Learning and performance are separate, with training done before performance. The algorithm designers not only design the learning algorithm, they also curate the dataset, which determines the direction learning progresses in. This is typical in many classification (including Deep Learning) systems. A standard practice in AI systems is to train on data and then perform with what has been learned.
   [Timing: External; Direction: External; Data: External; Learning: Internal]
2. Staggered performance and learning. There are AI architectures, such as Prodigy and FORR, where an agent performs a task, and then the learning component analyzes the agent's performance (often a trace of behavior). The learning component creates knowledge that is added to the agent. In the future, the agent uses that knowledge to improve performance.
   [Timing: External; Direction: External; Data: Internal; Learning: Internal]
3. On-line architectural learning. Some learning AI systems have automatic on-line learning components that continually run in the background while the system is performing the task. This is what we classify as Level 1. For example, almost by definition, reinforcement learning involves a component where there is on-line learning of the value function (such as Q-learning). Other examples may include underlying SLAM mechanisms in some robots. The learning mechanisms in Soar and ACT-R fit in this category. One of our activities is categorizing these L1 mechanisms across AI systems and cognitive architectures. Our analysis suggests that very few architectures support more than L1 mechanisms (such as L2 strategies) as learning architectures often because of a lack metareasoning capabilities, whereas many metareasoning systems lack architectural learning mechanisms.
   [Timing: Internal; Direction: External; Data: Internal; Learning: Internal]
4. On-line controlled learning. Some AI systems do not treat learning as an automatic process, but instead have task knowledge (written by a programmer)

that determines when and what to learn for each task. This approach is prevalent in AI agents developed directly in a programming language or in AI frameworks (such as Blackboard systems) that support agent development but do not have a fixed set of learning primitives. DIARC is an example where any module can include a code that stores knowledge during performance. Robotic architectures that deliberately update their maps and knowledge of the world (such as CoBots). Within architectures, there are a few examples. Soar agents can deliberately store information into semantic memory, although we are investigating how this can be replaced by an architectural learning mechanism. [Timing: Internal; Direction: External or Internal; Data: Internal; Learning: External]

On-line learning but without ongoing embedding in its environment. Here the learning system is not completely autonomous but instead there is external control (usually by a human) of aspects of the learning experience. This approach is prevalent in the RL community, where the AI system does on-line learning for a specific trial, but a human (or supervisory program) determines the *context* of the trial, including an ordering of trials, the initial situations, the rewards, and possibly even sweeps over hyper-parameters. Thus, the AI system *is not truly embedded in its environment* – it is more of a tool than part of an organism with an ongoing existence. Deep RL systems such as AlphaZero and AlphaStar fall under this category where they have structured curriculum training, and do not exist in a permanent environment. Others such as Go-Explore control the environment in ways outside of the agent's experience (jumping to specific states). One value in examining these systems is that we can identify activities currently performed by humans that might be valuable for a truly autonomous agent to perform for itself (if possible). There is a related class of ML methods that basically freeze various parts of the architectures after some learning experiences (example: https://arxiv.org/abs/1904.12584).
[Timing: Internal; Direction: External or Internal; Data: External; Learning: External or Internal]

5. On-line learning *supplemented* by external interaction. This is the case where an agent is embedded in an environment, but its experience can be supplemented with interactions with other agents. For example, in a classroom, the teacher is responsible for a child's experience, but the child is embedded in that environment, and the teacher cannot directly change that embedding, or influence its experience. Similarly, in some AI systems, live human-generated reward can be incorporated in RL algorithms to influence how learning progresses without influencing how the agent is embedded in its environment,

or live human-generated instruction can be interpreted by the agent as commands that it *should* obey, but the agent has control over its own actions.
[Timing: Internal; Direction: External; Data: Internal; Learning: External]

6. On-line architecture learning (L1) supplemented with internal metacognitive reasoning (L2) - deliberate strategies to manage learning experience by the agent. To date, we have found only rare cases of this in AI systems, such as in Rosie where it explicitly does a retrospective analysis to learn a policy after it has performed a task. Note that the difference between this and approaches described above is that the decision and processing involved in the retrospection is completely internal to Rosie – no human is involved.
[Timing: Internal; Direction: Internal; Data: Internal; Learning: Internal]

4. Cognitive Capabilities and their Relationship to Levels of Learning

One surprise is the dearth of Level 2 in AI agents and architectures. This led us to examine the characteristics of tasks, agents, and environments that lead to systems with Level 1 and Level 2 learning, and how those characteristics are missing in many current AI systems. Below are the characteristics we have identified that appear to be the most important.

a. Embodiment: Ongoing, long-term, embodied existence in a dynamic environment. Few AI systems take ongoing existence as a requirement and assume that humans are available to set up training situations, etc. Thus, many AI systems are treated as *kept* systems that are not autonomous. This makes it possible for external control of their embedding and experience and has made it possible to avoid confronting how an embodied agent can have more control over its learning experiences.
b. Metacognitive capabilities: The agent needs to be able to "step back" and create a learning goal – not just a task goal. Exactly which metareasoning capabilities are required is still unclear, but some seem necessary. A hypothesis is that one additional reason for the lack of research in Level 2 systems is because there are very few learning systems with metareasoning capabilities. Existing research on metacognitive capabilities for intelligent agents focus on goal-oriented decision making assuming the knowledge that underlies it is constant. It is not immediately clear that the proposed metacognitive reasoning frameworks can be applied to learning systems where knowledge evolves with experience.
c. Computational limits: The agent has some computational and temporal limitations in its processing. Our hypothesis is that the split between L1 and L2 arises in part because L1 mechanisms are designed to run in the background, so they do not interfere with an agent's task pursuits. In contrast, many types of L2 reasoning require much more extensive processing and

cannot fit within the primitive computational cycle of an agent. However, we also assume that there are situations where the agent is not under time pressure, so L2 mechanisms can be used.

d. Self-modeling: The agent has some model of its own L1 learning capabilities so it can plan activities that invoke them productively.
e. Episodic memory. The agent can refer back to prior experience and either replay it or mine it for important events and regularities that help build a model of its own learning.

Papers:

1. Mohan, S., Klenk, M., Shreve, M., Evans, K., Ang, A., & Maxwell, J. (2020). Characterizing an Analogical Concept Memory for Architectures Implementing the Common Model of Cognition. In the Proceedings of the Annual Conference of Advances in Cognitive Systems (invited to the journal).
2. Stearns, B. (2021). A Comprehensive Computational Model of PRIMs Theory for Task-Independent Procedural Learning. Ph.D. Thesis, University of Michigan.