

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

John E. Laird
University of Michigan
2260 Hayward Street
Ann Arbor, MI 48109-2121
laird@umich.edu

Shiwali Mohan
PARC
3333 Coyote Hill Rd.
Palo Alto, CA 94304
shiwali.mohan@parc.com

Abstract

We propose two distinct levels of learning for general autonomous intelligent agents. Level 1 consists of fixed architectural learning mechanisms that are innate and automatic. Level 2 consists of deliberate learning strategies that are controlled by the agent's knowledge. We describe these levels and provide an example of their use in a task-learning agent. We also explore other potential levels and discuss the implications of this view of learning for the design of autonomous agents.

Introduction

General autonomous intelligent agents have many challenges when it comes to learning. They must continually react to their environment, focusing their computational resources on making the best decision for the current situation using all their available knowledge. They also need to learn everything they can from their experience, building up their knowledge so that they are prepared for making the best decisions in the future. Learning must be ubiquitous, on line, and not disruptive when task demands are high. However, an intelligent agent can also step back and deliberately structure its behavior to enhance its learning. It can pursue tasks whose purpose is to acquire knowledge, such as through rehearsal, imitation, retrospection, reading a book, attending a lecture, or even pursuing academic research.

In this paper, we propose that in human-like agents, learning can be split into two levels, inspired by research across neuroscience, psychology, and cognitive architecture (Laird, Lebiere, and Rosenbloom 2017). Level 1, or L1, encompasses *architectural learning mechanisms* that are innate, automatic, on line, and effortless, such as the temporal difference update mechanism in reinforcement learning. Level 2, or L2, encompasses *deliberate learning strategies* that are realized through knowledge and controlled by an agent. These strategies are essentially tasks that the agent adopts, and they compete with other tasks for mental (and physical) resources. In addition, they themselves can be learned. A simple example in humans is deciding to explicitly rehearse a phone number in order to memorize it. Deliberately repeating the number several times aloud (or to one's self)

creates the experiences that are consolidated by automatic L1 memory mechanisms, making the number available for later recall. L2 strategies do not include additional learning mechanisms, but instead take advantage of the underlying L1 mechanisms to extract regularities and record knowledge structures from the generated experiences.

We posit that these two learning levels are ubiquitous in humans and provide a useful dichotomy for thinking about and developing learning in long-lived general autonomous intelligent agents. L1 mechanisms learn no matter what time or task constrains there are. L2 strategies come in when the agent has time to deliberately attempt to improve its learning. An intriguing hypothesis that L2 methods are unique to humans. We have the meta-cognitive capabilities to both reflect on our behavior and learning, develop strategies for improving them, communicate them to others, and use them. Animals appear to be missing many or all these capabilities, at least in the depth and breadth possible in humans, and as a result they do not deliberately create or use strategies whose goal is to improve their learning.

In this paper, we examine these levels, provide examples, and discuss their implications for agent design, development, and long-term existence. We describe some of the processing stages found in Level 2, and illustrate them with an agent we have developed that can learn completely new tasks from natural language interaction. We then explore the possibility of levels in addition to L1 and L2 and conclude.

Level 1: Architectural Learning Algorithms

L1 consists of architectural learning algorithms that automatically and continually extract regularities from an agent's experience and reasoning, and directly modify the agent's long-term memories. They are innate, fast, effortless, and outside the agent's control. We cannot explicitly invoke them ("I will learn this right now!") nor can we explicitly inhibit them ("I refuse to learn this!"); however, as described below in Level 2, we can adopt *strategies* to influence what they learn. From a cognitive architecture perspective (Laird, Lebiere, and Rosenbloom 2017), they include learning mechanisms for all of the types of long-term memories found in a general agent: perceptual, procedural, motor, episodic, and semantic. From a psychology perspective, L1 is related to *implicit learning* (Ebbinghaus 1885), where the subject is not aware that *they are learning* (Frensch and

Runger 2003). Many definitions emphasize not being aware of *what is learned*, but that is outside of our concerns. Examples of L1 learning from behavioral psychology include operant conditioning, classical conditioning, habituation, sensitization, and rote learning.

There are no restrictions on the types of knowledge representations that an L1 algorithm can learn over. They can learn directly from an agent's perceptual stream, but also simple feature-based statistical representations and internally created relational symbolic representations. We expect that a general autonomous agent would have many different L1 learning mechanisms, each optimized for different types of data. Some L1 algorithms could learn from short-term, high frequency changes, while others integrate data over longer time scales. L1 mechanisms are characterized by the way they are embedded within an agent's processing. If a mechanism is innate and automatic, then it is L1.

Level 2: Deliberate Learning Strategies

L2 consists of deliberate learning strategies that create the experiences from which L1 algorithms learn. When a student decides to write word-pairs on two sides of index cards and then train and test themselves, that is an L2 strategy. When an athlete decides to practice their three-point shot to raise their shooting percentage, that is an L2 strategy. L2 strategies are voluntary, deliberately initiated by agent reasoning and knowledge, becoming a goal or task that directs behavior. In pursuit of them, an agent can use any and all of its cognitive capabilities, such as, analogy, attention, decision making, dialog, goal-based reasoning, meta-reasoning, natural language reasoning, planning, spatial reasoning, temporal reasoning, and theory of mind to generate the experiences from which the L1 mechanisms learn. Given their deliberate nature, they compete with other agent tasks, and whereas an L1 strategy is automatic and continuously active, an L2 strategy can be interrupted by another task and will be pursued only when there are not more urgent competing tasks. Thus, an L2 strategy allows an agent to use complex reasoning for learning that is not possible with L1, but with the trade off that the L2 strategy can be used only when the agent is not pursuing some other task (unless the agent can intermix the L2 strategy with the other task).

In contrast to L1 algorithms, which are prisoners to the agent's ongoing experience, an L2 strategy has the ability to control the agent's experiences. An L2 strategy can determine which problems to solve (deliberate training or exploration), recall and analyze past experiences (retrospective analysis and combining temporally distance information), imagine future or hypothetical situations (prospective analysis), deliberately map a problem onto a similar previously solved problem (analogy), or interact with another agent (learn by demonstration or instruction).

Note that these are *learning* strategies, meant to enhance long-term performance. They are not the same as short-term *information gathering* strategies, whose purpose is to gather additional data that aids in solving the agent's current problem(s). Such strategies may indirectly enhance learning (by creating new experiences that L1 mechanisms learn from), but that is a byproduct of those tasks, and not their goal.

Some L2 strategies can be learned through reflection over one's own behavior, such as noticing that internal rehearsal enhances one's ability to recall a phone number in the future, or that practice on a physical task improves performance. Other L2 strategies can be learned by watching what others do, or by being taught to use a certain an L2 strategy without really knowing why. Many of the most effective L2 strategies are less intuitive because they take advantage of subtle strengths and weaknesses of L1 mechanisms to improve learning. These strategies are discovered through deliberate observation and experimentation, such as performed in educational psychology (Brown, Roediger, and McDaniel 2014). Below are six such L2 strategies described by Weinstein, Smith, and Caviglioli (2016).

1. Spaced Practice: "Space out your studying over time."
2. Retrieval Practice: "Practice bringing information to mind without the help of materials."
3. Elaboration: "Explain and describe ideas with many details."
4. Interleaving: "Switch between ideas while you study."
5. Concrete Examples: "Use specific examples to understand abstract ideas."
6. Dual Coding: "Combine words and visuals."

In AI, Level 2 learning strategies are related to meta-learning (Briggs 1985), which in one formulation is "being aware of and taking control of one's own learning" (Maudsley 1979). These formulations do not relate deliberate learning to the architectural learning mechanisms described here.

Level 2 Learning in a Task Learning Agent

The concept of different levels of learning is not new. Cognitive architectures such as ACT-R and Soar (Laird 2012) subscribe to the distinction between architectural and deliberate learning, as does a proposed Standard Model of the Mind (Laird, Lebiere, and Rosenbloom 2017). Both ACT-R and Soar have multiple automatic architectural learning mechanisms: rules are created, activation and utility values are adjusted, and declarative memory structures are created automatically according to fixed algorithms. In Soar, agent knowledge can directly store knowledge into semantic memory, but we see that as a temporary anomaly, and are actively researching architectural learning mechanisms to replace it.

In both Soar and ACT-R, agent behavior indirectly determines what is learned by the underlying architectural learn mechanisms, and L2 strategies are encoded as procedural and declarative long-term knowledge. Although there can be a wide variety of L2 strategies in an agent, below we outline four phases of processing that are characteristics of L2 strategies. These phases include: identifying that there is a learning opportunity, selecting a learning strategy, executing the strategy, and then using the learned knowledge.

We illustrate these phases using an existing robotic agent, called Rosie (Mohan and Laird 2014), implemented in Soar. Rosie is one of new breed of Interactive Task Learning (ITL) agents (Laird et al. 2017) that learn new tasks from scratch through natural interaction with a human instructor.

Rosie learns over thirty different puzzles and games (Kirk and Laird 2016), as well as navigation and delivery tasks (Mininger and Laird 2018). Rosie does not have a separate task learning module or component, but learns the definition of a task through Soar’s fixed L1 learning mechanisms, combined with its knowledge for processing and interpreting natural language instructions.

Identifying a learning opportunity

Reasoning about applying an L2 strategy begins with identifying an opportunity where new knowledge can be acquired. Information to recognize this opportunity can come from multiple sources:

1. **Missing, Conflicting, or Uncertain Knowledge:** The agent may detect that the knowledge needed for its current task performance is inadequate, inconsistent, or too uncertain to apply.
2. **Unexpected Event or Novel Situation:** The agent detects an unexpected event in its environment, or novel situation (new terrain, new objects, and novel configurations), suggesting its model of the world is incomplete.
3. **External Knowledge:** The agent recognizes that an external source of knowledge is available to learn from. This could be an observer who explicitly identifies that the agent’s knowledge is lacking, or it could be a passive source, such as a book. It could also include observing behavior of another agent that achieves a desired goal such as learning by imitation or observation.
4. **Past Experiences:** The agent has knowledge that certain activities (such as practice, study, or research) can enhance its future performance, and gaining that knowledge is more important than other, competing tasks.

Rosie relies on the first two cases, using Soar’s automatic ability to detect when knowledge is missing or conflicting. In those situations, Soar generates a subgoal in which Rosie can deliberately attempt to derive or discover the additional knowledge needed to allow problem solving to proceed. These cases arise in Rosie in many situations, including when it does not know the definition of a task component, such as the goal, or when its own reasoning fails to determine which action to take.

Selecting a learning strategy

Once a learning opportunity has been identified, the agent must choose how to respond. This includes picking an appropriate L2 strategy, but it also involves deciding that attempting to learn is better than alternative actions. In some situations, there might not be time to employ a learning strategy and it is better for the agent to “plow” ahead, choosing a task-relevant action even if it is not completely confident in it. The exploitation/exploration trade-off in RL is one example of this. These decisions can be based on a combination of general heuristics the agent is preprogrammed with, but also situation-specific knowledge the agent learns a key aspect of L2 methods is that they can themselves be improved through learning.

Currently, Rosie uses the single-minded strategy to always attempt to acquire missing task definition knowledge through interaction with the instructor. It does not reason about what is the best way to learn. Once it has learned a task definition, but does not know the best way to pursue it, Rosie attempts an internal search to discover a solution. If it is unable to find a solution on its own, either because of an insufficient world model, or because internal search is too deep, it will ask for advice from the instructor.

Executing the learning strategy

Once an L2 learning strategy is selected, the agent executes it. For some strategies, this may involve a retrospective analysis of previous experiences, retrieved from episodic memory, while for others, it may involve explicitly setting up different world situations for the agent to experience and experiment with. These are deliberate strategies, and they are always competing with other activities the agent can consider, so that at any time, the agent can decide to abandon a learning strategy and pursue other actions if it determines those actions are more important than learning.

If Rosie required instruction to correctly perform a task, later, that experience is recalled from episodic memory as a part of retrospective analysis. Rosie replays the experience while performing a causal analysis of why that sequence of actions and instructions leads to goal success (Mohan and Laird 2014). Currently, Rosie does not reason about when is the correct time to apply this learning strategy. It always applies it right after successful task execution that was driven by instruction. In more realistic scenarios, there can be other task pressures such that the agent must delay reflection and retrospection to a later time.

Using learned knowledge

Once new knowledge has been acquired, it is available to be used. A final deliberation could be to test, verify, and monitor acquired knowledge to ensure its correctness. This can be guided by the agent who identifies other similar tasks, applies what is learned, and evaluates its behavior. During this process, the agent may identify missing knowledge or inconsistencies and consequently, start phase 1 over again. Or, an intelligent observer can identify other tasks that may be useful for evaluating the agent’s learned knowledge.

In Soar, and thus Rosie, once knowledge has been acquired, it is immediately available for use by the agent for any future problem. Knowledge can transfer to related problems, and in our work with Rosie we have demonstrated that as Rosie learns more tasks, the time to learn additional tasks decreases significantly whenever there are overlapping concepts, subgoals, or behaviors.

Beyond L1 and L2

Our focus with L1 and L2 has been on distinctions between different types of learning mechanisms and strategies used by an agent during its extended existence. If we broaden our analysis beyond learning within a single agent, we can identify additional levels of learning.

Level 0: Population evolution

We posit that L2 mechanisms are learned through L1 processes. Where do L1 mechanisms come from? It seems clear that in humans and animals, it is the evolutionary process that creates the L1 mechanisms within our brains. For artificial agents, it is the human designers (at least until the singularity hits) that develop the L1 mechanisms. In natural systems, there might also be a development level (0.5?) that refines or specializes innate L1 mechanisms.

Level 1+: Innate strategies

In level 2, the agent's knowledge is used to decide on behavior whose goal is to learn. In natural (and artificial) systems, there can also be innate drives that direct behavior, creating experiences that aid learning. However, we do not classify them as L2 strategies because they are not under the agent's control. These L1+ mechanisms can be driven by intrinsic motivators, such as curiosity. One example is the tendency for young animals to play, which helps them learn. But we claim it is a stretch to say that they have an explicit goal to learn. Another example is imitation, where there appears to be an innate desire to imitate, which can lead to learning new skills. Imitation can be an effective L2 strategy when it is *deliberately* selected with the purpose of learning, such as in learning by demonstration or observation, but if it is initiated because some innate pleasure results from the imitation, we do not classify it as an L2 strategy.

Level 2+: Social strategies

The levels we have defined are from the perspective of an individual agent. In social agents, the initiation of learning experiences can be driven by other agents (teachers and mentors), where the original agent relies only on its L1 capabilities, but obtains the benefits of the teaching and mentoring abilities of other agents. Our society depends on this multi-agent structure to enhance learning, both informally in learning from immediate family members and more formally through our education system. We resist including this as a completely different level as it is not a property of the individual agent. However, it is likely that many L2 strategies are tightly bound to social interaction. L2 might be necessary for maximally successful learning in social situations, and social interaction may play a significant role in an agent's acquisition and use of many of its L2 strategies.

Level 3: Deliberate strategies for modifying L1

For completeness, we include level 3 for behaviors that attempt to modify, either directly or indirectly, L1 mechanisms. In humans, these involve behaviors that attempt to change the underlying physiology of learning mechanisms (in the brain) to make them more effective. Some examples are getting plenty of rest, exercising, and ingesting cognitive enhancing drugs (nootropics) that improve concentration and attention (such as stimulants). It is less clear how this level currently applies to artificial agents, although there have been examples in science fiction, such as in the *Westworld* TV show, when the Maeve Millay (AI) character gains direct access to her underlying cognitive capabilities and deliberately increases her intelligence.

Discussion

The title of this paper is inspired by Daniel Kahneman's book, "Thinking Fast and Slow," where he explores the distinction between two levels of thinking (Kahneman 2011). The first, labeled System 1, is uncontrolled, associative, and "intuitive" thinking, corresponding most closely to reactive systems in AI. The second, System 2, is slower and supports deliberate, logical, and more rational thinking. This distinction was originally identified by William James (James 1890) and has had a long and storied history of research under dual process theory (Groves and Thompson 1970). Here we've extended the idea of two levels to learning, with Level 1 sharing the automatic and uncontrolled aspects of System 1, and Level 2 relying on System 2 thinking to generate experiences for Level 1 learning.

The mapping of levels of learning onto fast and slow appears to be a bit more complicated than it is for thinking. Many L1 mechanisms must be fast in processing the data that streams through the agent. However, some L1 mechanisms may be slower, detecting regularities that are extracted from data that persists across longer time scales, maintained by the agent's reasoning as the agent works on a protracted problem. And what about what is learned? We know that humans (and animals) can learn some things exceedingly fast, often in one shot, seemingly depending on only L1 mechanisms. In contrast, many current L1 mechanisms studied in AI and ML are excruciatingly slow, requiring thousands of trials to learn new policies or concepts. However, we must be careful in making comparisons because these AI agents are often learning from scratch, whereas the humans are building on large bodies of existing knowledge.

In comparison to L1 mechanisms, the *processing* for an L2 strategy is slow, requiring extended time (often minutes to hours) to marshal knowledge and generate the appropriate experiences. For example, learning from instruction requires minutes of interactions, not milliseconds to acquire new knowledge. However, one of the strengths of L2 strategies is that they combine knowledge from multiple sources, so that meaningful concepts can be learned fast, often in one shot. Moreover, they can combine knowledge learned from different sources, explicitly generalizing that knowledge so that it can apply to situations the agent has never experienced (so called "zero shot" learning).

In terms of fast and slow, what is really important in our analysis is the combination of L1 within L2. There are ways in which each can be slow, supporting incremental, cumulative learning over time, but together, they can also be incredibly fast. For example, in one viewing of a TV show, we can learn detailed information about characters, their personalities and quirks, their relationships with other characters, and the sequences of events that immerse them in a complex plot in which a "game is afoot." All of which are available both immediately and in the future, for recall, further analysis, and endless discussion (Baldassano et al. 2017).

Finally, what is the impact of this analysis on how we design autonomous agents? There are many possible ways to integrate learning into general autonomous agents. Throughout the history of AI, there has been a tendency to develop independent learning mechanisms, especially for complex

learning, such as learning by analogy, learning by instruction, learning by demonstration, learning by induction, inverse reinforcement learning, and so on. Often these are studied in isolation, and are built with special-purpose learning mechanisms. In contrast, the approach we propose here is different and possibly radical. It suggests that when trying to develop general agents with broad learning capabilities, it is possible (not necessarily necessary, but possibly sufficient) to develop a core set of primitive, automatic learning mechanisms that are shared by complex deliberate learning strategies mechanisms. The deliberate strategies leverage these primitive mechanisms, and do not have any strategy-specific learning mechanism of their own. We see this as an exciting path forward for the development of general autonomous intelligent agents.

Acknowledgments

The ideas in this paper sprung from research on learning in cognitive architecture with graduate students in the Soar lab at the University of Michigan, and from many conversations over the years with Allen Newell and Paul Rosenbloom.

The work described here was supported in part by the Air Force Office of Scientific Research under Grant Number FA9550-15-1-0157. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the AFOSR or the U.S. Government.

References

- Baldassano, C.; Chen, J.; Zadbood, A.; Pillow, Jonathan, W. P.; Hasson, U.; and Norman, K. A. 2017. Discovering event structure in continuous narrative perception and memory. *Neuron* 95(3):709–721.
- Briggs, J. B. 1985. The role of meta-learning in study process. *British Journal of Educational Psychology* 55:185–212.
- Brown, P. C.; Roediger, H. L.; and McDaniel, M. A. 2014. *Make it stick, The Science of Successful Learning*. Harvard University Press.
- Ebbinghaus, H. 1885. *ber das Gedchtnis. Untersuchungen zur experimentellen Psychologie [Memory: A Contribution to Experimental Psychology]*. Leipzig, Germany: Duncker & Humblot.
- Frensch, P. A., and Runger, D. 2003. Implicit learning. *Current Directions in Psychological Science* 12(1):13–18.
- Groves, P. M., and Thompson, R. F. 1970. Habituation: a dual-process theory. *Psychological review* 77(5):419.
- James, W. 1890. *The principles of psychology*. Henry Holt and Company.
- 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.
- 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.
- 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. E. 2012. *The Soar cognitive architecture*. MIT press.
- Maudsley, D. 1979. A theory of meta-learning and principles of facilitation: An organismic perspective. Technical Report (40, 8,4354-4355-A), University of Toronto.
- Mininger, A., and Laird, J. E. 2018. Interactively learning a blend of goal-based and procedural tasks. In *AAAI*.
- Mohan, S., and Laird, J. E. 2014. Learning goal-oriented hierarchical tasks from situated interactive instruction. In *AAAI*, 387–394.
- Weinstein, Y.; Smith, M.; and Caviglioli, O. 2016. Six strategies for effective learning. <http://www.learningscientists.org/blog/2016/8/18-1> [Accessed: November 21, 2017].