Animal Learning and Environment Models

Andrew Perrault
perrault.17@osu.edu
CSE 5539

Recap: intelligence

Starting point: how to build artificial intelligence?
- Defined intelligence as task capability
- As in humans, intelligence is multidimensional in machines
- Narrowly intelligent systems are intelligent in only a few closely linked tasks (e.g., AlphaFold, AlphaGo, Stable Diffusion)
- Generally intelligent systems are intelligent in many tasks (e.g., LLM/MLLM-based systems that can take any task as input)
- Task-specific pattern recognition is the main driver of human performance in tasks, and can be improved through practice
- No “far transfer” in humans: task X patterns don’t transfer to task Y unless the tasks are very close (e.g., French → Spanish). Thus practicing task X does not improve task Y performance (commonly investigated in chess, music, memory training)

Recap: building via imitation

Idea #1: imitate known intelligence (e.g., humans)
- We have developed powerful imitative systems (fast digital computers and “the transformers recipe”) over the last 70–200 years
- The imitative approach has yielded many useful systems (e.g., AlphaFold, base LLMs, Stable Diffusion)
- Imitation could be limiting: hard to surpass best current task behavior, strong performance data may be scarce, machine capabilities may be different from the system they are imitating (and imitative systems believe they have the capabilities of the source of imitation data)

Recap: learning from experience

Idea #2: learn from task experience (reinforcement learning)
- System performs task directly, learn from successes and failures
- *Advantage*: specify desired **outcome**, not desired **behavior**
- *Advantage*: does not require high quality data (and, thus, performance not limited by data quality)
- *Disadvantage*: building an environment model might not be easy
- *Disadvantage*: **learning without a teacher is much harder**

Recap: learning without a teacher

Worst case: too many actions to try, no hope of learning (alien box example)

But everything humans have learned has been without a teacher

Hope that computer-based systems, with their different capabilities, could broaden human understanding

Recap: making RL easier

Because of the difficulty of learning in large action spaces, several strategies for reducing the space of actions we must try:
- Pick tasks where we can try almost everything (board games, Atari)
- Use mathematics to reduce search space (control theory, robotics)
- Use an imitative prior (LLMs/MLLMs)

We saw an example of the imitative prior + RL pipeline in GPT-3
- Imitative prior produces grammatical completions with limited meaning
- After RL, collapse to a small number of completions that have meaningful semantic structure

Imitative prior + RL pipeline can outperform best single expert by “stitching” experts together
- But still limited by what prior includes

This lecture: learning from animal behavior

How do animals learn?

What is the difference between pure trial-and-error and planning in an environment model?

Philosophy of animals

Aristotle: animals above plants and below humans, have only basic desires

Rene Descartes (early 17th c.): animals have no minds, no souls, no true feelings

Margaret Cavendish (1650s-60s): animal behavior is too complex to be mindless
- Examples: spider's web, bird's nest

Darwin (mid 19th c): there is no hard line between humans and animals
- They share a common ancestor via evolution
- *The Expression of the Emotions in Man and Animals*: same basic emotions are present across species, i.e., emotions are evolved, adaptive traits
- Consequence: *comparative psychology*—we can study animals to understand humans

Can a cat escape a puzzle box (for fish)?

Edward Thorndike (1890s): we can understand *learning* in humans via learning in animals

Task: a cat in a puzzle box has to do a specific action to escape (e.g., pull a lever)
- If it escapes, it gets a fish!
- You may recognize this experiment from the intro

Mechanism for escape is hidden
- The cat can't just study the problem and decide what to do
- It has to experiment with random stuff (touching everything, meowing, biting)

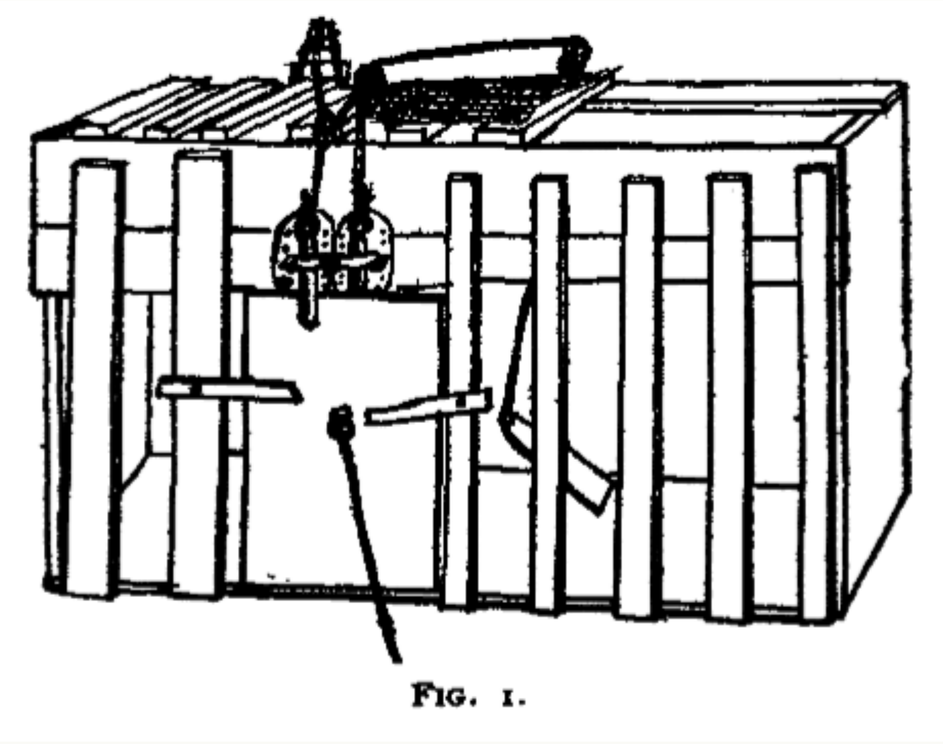
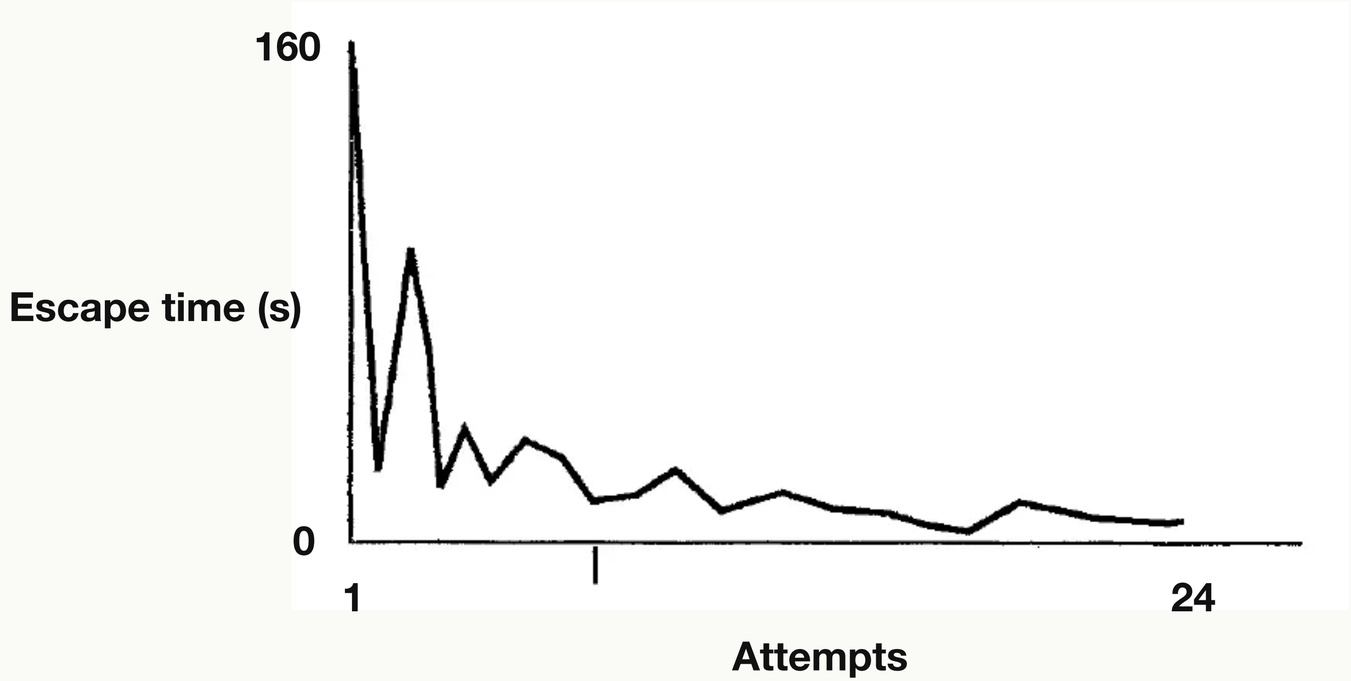


FIG. 1.

Subject: Cat 12, Task: Box A

Box A: pull a hidden loop

Very long initial escape time of 160 s → gradual improvement to 6s

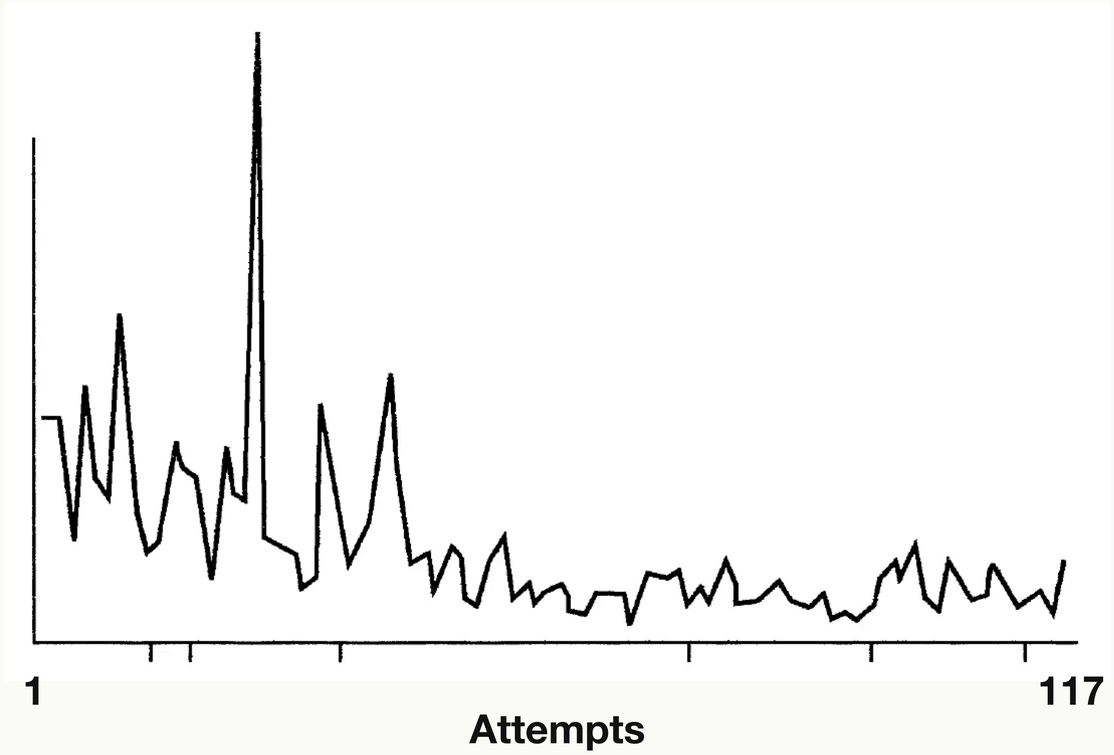


Thorndike, 1898

Subject: Cat 4, Task: Box K

Box K: three physical mechanisms that must be manipulated in sequence (lever → string → bar)

Slower and less stable learning



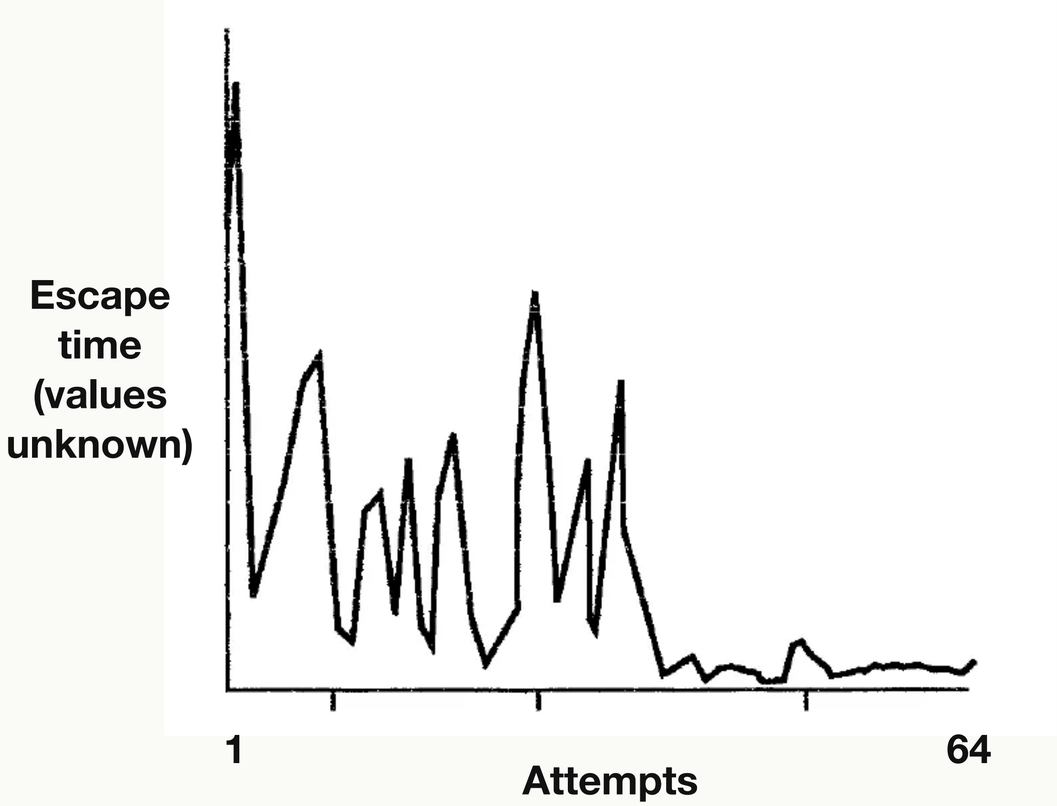
Thorndike, 1898

Subject: Cat 5, Task: Box Z

Box Z: lick yourself and the experimenter will open the door

Really hard to learn, very bouncy

Cat has prior that physical obstruction should be opened by physical interaction?



The “Law of Effect”

Result: cats get *gradually* faster at escaping over time
- Behaviors that lead to a reward are “stamped in” (aka reinforced)
- Behaviors that lead to nothing are “stamped out”

Thorndike’s conclusion: humans learn via trial-and-error
- Human learning is not “about” understanding, it’s about trying random stuff until it works
- Reason is a network of habits (patterns): a smart person is someone who, through trial-and-error, has built up a larger library of correct response

Can a chimpanzee eat unreachable bananas?

Wolfgang Köhler (1910s): learning in “higher animals” is not just trial-and-error, they are capable of *insight*
- Thorndike’s experiments are “rigged” because cats can’t see the mechanism

Task: banana is hung out of reach.
- Bananas are delicious
- Boxes, sticks, etc. available—tools that can be used to reach the banana

What do you think happened? Did they learn gradually, like cats in Thorndike’s experiment?

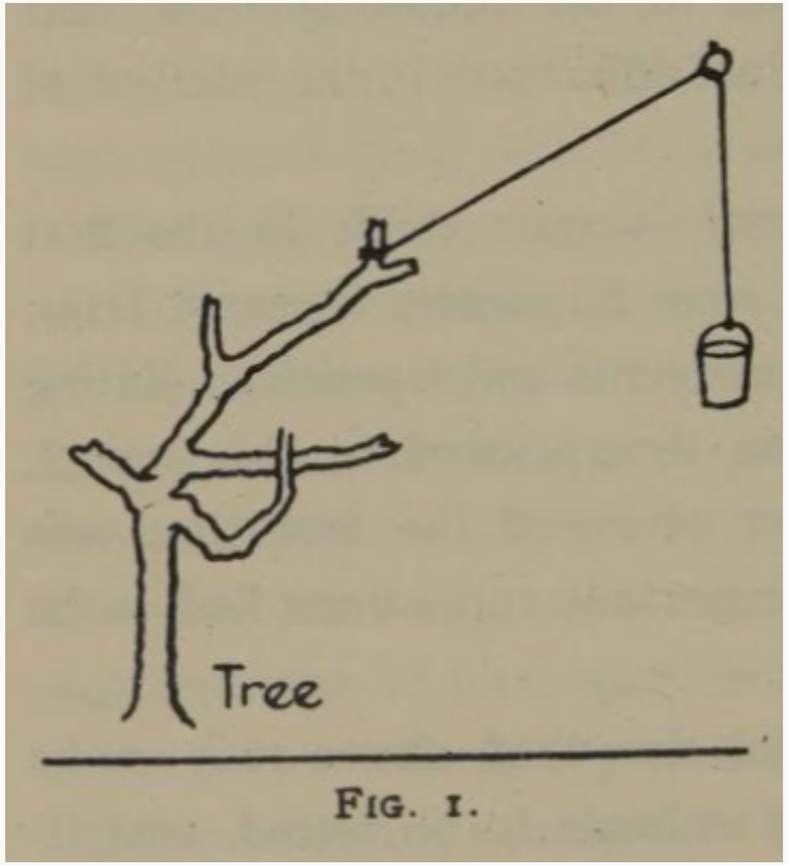


FIG. 1.

Video

https://www.youtube.com/watch?v=6-YWrPzsmEE

Insight learning

First, Sultan tries jumping, gets frustrated. Then, he'd have an "aha" moment and use a tool to get the bananas

Köhler: Thorndike was wrong
- Human learning is not about trial-and-error
- We see the whole picture of a problem, "mentally restructure" the elements, and the solution appears to us suddenly, fully formed

PLATE III. SULTAN MAKING A DOUBLE-STICK

Is insight learning RL?

In insight learning, the agent does not learn by interacting with the environment

But what is happening during those 10 minutes?

The agent must be thinking about how to solve the problem

Thinking about how to solve the problem = trying different stuff in an environment model

And now (slippery slope), any solution method for an environment is RL ■

Köhler’s criticism of Thorndike

Why did Köhler say that Thorndike’s experiments are rigged because cats can’t see the mechanism?
- I.e., why did that matter?

Köhler: **nobody** could plan in Thorndike’s experiments because the mechanism is *hidden*
- To be able to plan, a useful model of the environment has to be accessible to you
- Thorndike’s experiment was a poor fit for his question of “how do people/cats learn?”

Can cats plan?

Yes, e.g., they can simulate the path of a mouse they are chasing
- But they don’t understand tools, e.g., string, sticks, boxes
- When they can’t represent something, they default to trial and error

Cats Chasing Mouse Funny Compilation by Somm37

Can chimps trial-and-error?

Yes, Sultan tried a bunch of stuff before coming up with his insight

This process is probably needed to create an accurate environment model to think about
- E.g., are the boxes too heavy to lift? Is the stick strong enough?

PLATE II. CHICA ON THE JUMPING-STICK, RANA WATCHING

Model-free vs. model-based RL

If an RL system plans using an environment model, that system is *model-based*

If there is no environment model used, that system is *model-free*

Insight learning (i.e., rapid improvements without interacting with the environment) only happen when planning based on an environment model

Thorndike cats: model-free approach to escaping puzzle boxes vs.
Köhler's chimps: model-based approach to eating unreachable bananas

Where are models used and not used in RL?

Model-based approaches dominate in almost Newtonian physics-based environment (e.g., robotics)
- We understand Newtonian physics very well and physics is relatively simple and consistent
- We don’t often need to experiment at all—just optimal control theory

Model-free approaches dominate in LLM RL
- We don’t have a good understanding of reasoning patterns
- Many model-based approaches have been tried in LLM RL, none have worked (yet)

Creativity

It **may** be that our definition of creativity is linked to model-based planning

Why?
- If an idea comes from pure trial and error, we may be reluctant to attribute creativity to it—we might call it **lucky**
- Creativity may require creative **intent** expressed by planning
- Perhaps **intent** requires a **goal** and a **goal** requires a **plan**

There are no current LLM outputs that are widely agreed to be creative
- Unlike AlphaGo's move 37, which clearly the result of planning (less than 1/10,000 probability before planning)
- This may change, e.g., work on, say, bin packing lower bounds (FunSearch), may be considered creative

Questions?